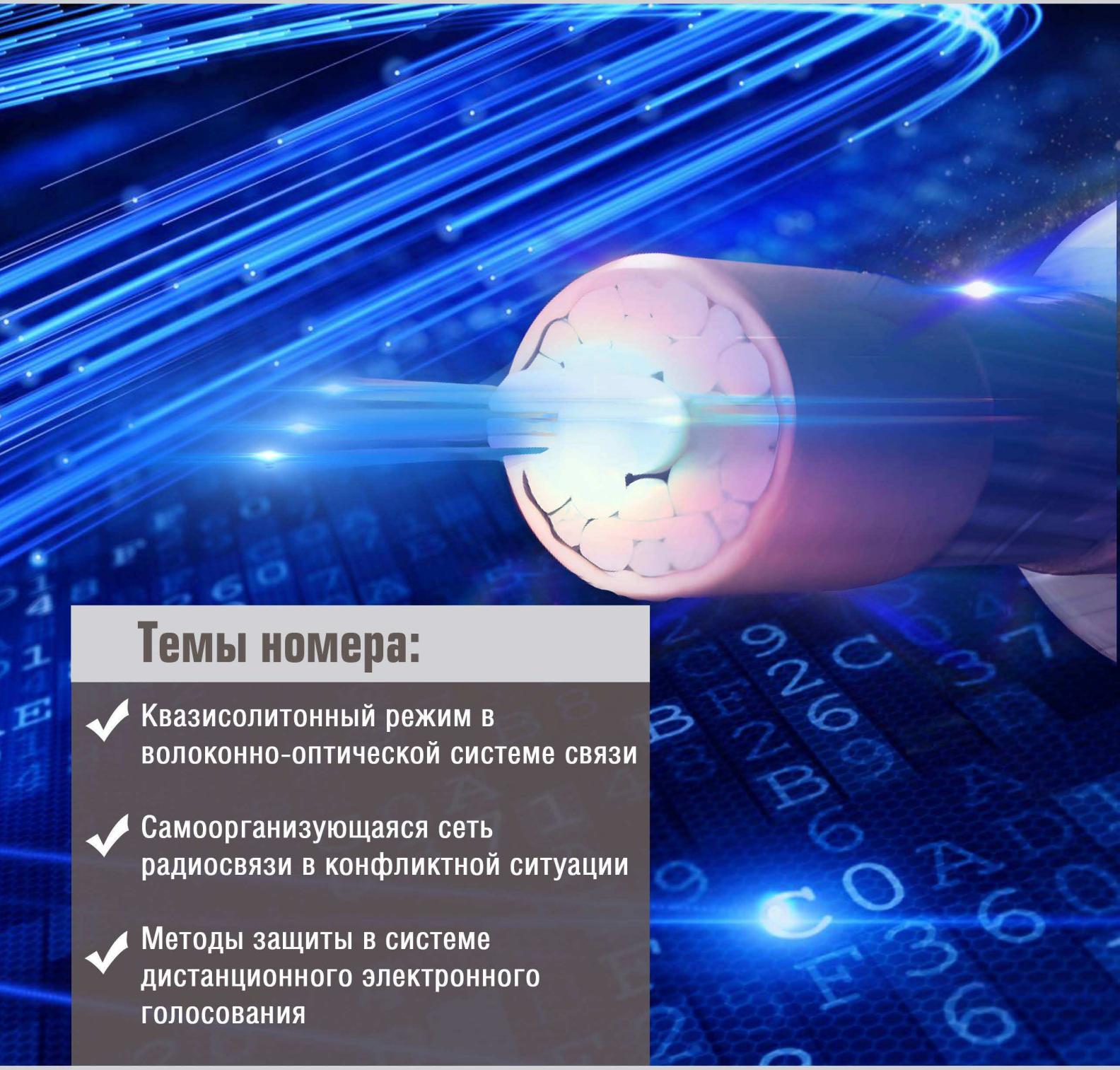
ISSN: 1813-324X (print)
2712-8830 (online)

ТРУДЫ УЧЕБНЫХ ЗАВЕДЕНИЙ СВЯЗИ

Том 9. № 2
2023



Темы номера:

- Квазисолитонный режим в волоконно-оптической системе связи
- Самоорганизующаяся сеть радиосвязи в конфликтной ситуации
- Методы защиты в системе дистанционного электронного голосования

Vol. 9. Iss. 2
2023

PROCEEDINGS OF TELECOMMUNICATION UNIVERSITIES

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

Научный журнал

ТРУДЫ
УЧЕБНЫХ ЗАВЕДЕНИЙ СВЯЗИ

Том 9. № 2

Proceedings of Telecommunication Universities

Vol. 9. Iss. 2

Санкт-Петербург

2023

Труды учебных заведений связи. 2023. Т. 9. № 2. ISSN: 1813-324X (print), 2712-8830 (online)

Описание журнала

Научный журнал. Включен в Перечень рецензируемых научных изданий, в которых должны быть опубликованы основные научные результаты диссертаций на соискание ученой степени кандидата наук, на соискание ученой степени доктора наук (распоряжение Минобрнауки России № 21-р от 12.02.2019), по специальностям (распоряжение № 33-р от 01.02.2022):

1.2.2. Математическое моделирование, численные методы и комплексы программ
2.2.6. Оптические и оптико-электронные приборы и комплексы
2.2.13. Радиотехника, в том числе системы и устройства телевидения
2.2.14. Антенны, СВЧ-устройства и их технологии
2.2.15. Системы, сети и устройства телекоммуникаций
2.2.16. Радиолокация и радионавигация
2.3.1. Системный анализ, управление и обработка информации
2.3.6. Методы и системы защиты информации, информационная безопасность

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

Редакционный совет

Регистрационная информация

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

Подписной индекс в объединенном каталоге «ПРЕССА РОССИИ»: 59983

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

Контактная информация

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

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

Адрес редакции: 193232, Санкт-Петербург, пр. Большевиков, 22/1, к. 334/2

Тел.: +7 (812) 326-31-63, м. т. 2022, +79643759970

E-mail: tuzs@sut.ru

Web: http://tuzs.sut.ru

ВК: http://vk.com/spbtuzs

Proc. of Telecom. Universities. 2023. Vol. 9. Iss. 2. ISSN: 1813-324X (print), 2712-8830 (online)

Description

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

1.2.2. Mathematical modeling, numerical methods and complexes of programs
2.2.6. Optical and optoelectronic devices and complexes
2.2.13. Radio engineering, including television systems and devices
2.2.14. Antennas, microwave devices and its technologies
2.2.15. Systems, networks and telecommunication devices
2.2.16. Radiolocation and radio navigation
2.3.1. System analysis, management and information processing
2.3.6. Methods and systems of information security, cybersecurity

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

Editorial Board

Registration Information

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

Subscription index for joint catalog «PRESSA ROSSII»: 59983

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

Contact Information

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

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

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

Phone: +7 (812) 326-31-63, local 2022, +79643759970

E-mail: tuzs@sut.ru

Web: http://tuzs.sut.ru

СОДЕРЖАНИЕ

CONTENTS

DOI:10.31854/1813-324X-2023-9-2

Научная статья
УДК 621.3.011.1
DOI:10.31854/1813-324X-2023-9-2-6-21

Математическая модель несимметричного вибратора с вынесенной точкой питания. Часть 2. Определение комплексной емкости малых, по сравнению с длиной волны, нормально разомкнутых проволочных антенн

Олег Вениаминович Попов[1] ✉, ov.popov@mail.ru
Андрей Витальевич Тумашов[1], ice47reg@yandex.ru
Георгий Николаевич Борисов[1], georgiiborisov@gmail.com
Константин Олегович Коровин[2], korovin.ko@sut.ru

[1]ООО «Специальный Технологический Центр»,
Санкт-Петербург, 195220, Российская Федерация
[2]Санкт-Петербургский государственный университет телекоммуникаций им. М.А. Бонч-Бруевича,
Санкт-Петербург, 193232, Российская Федерация

Аннотация: *Предложена методика вычисления комплексной емкости малых, по сравнению с длиной волны, нормально разомкнутых проволочных антенн. Получены аналитические выражения, определяющие взаимный потенциальный коэффициент двух любых произвольно расположенных линейных проводников. Предлагаемая методика может быть полезна при оценке потерь нормально разомкнутых проволочных антенн, размещаемых в непосредственной близости от полупроводящей поверхности, а также при определении их входного сопротивления.*

Ключевые слова: *комплексная емкость, взаимный потенциальный коэффициент, несимметричный вибратор*

Ссылка для цитирования: Попов О.В., Тумашов А.В., Борисов Г.Н., Коровин К.О. Математическая модель несимметричного вибратора с вынесенной точкой питания. Часть 2. Определение комплексной емкости малых, по сравнению с длиной волны, нормально разомкнутых проволочных антенн // Труды учебных заведений связи. 2023. Т. 9. № 2. С. 6–21. DOI:10.31854/1813-324X-2023-9-2-6-21

Mathematical Model of the Unbalanced Monopole Feed. Part 2. Determination of Complex Capacitance of Normally Open Wire Antennas, Small with Respect to the Wavelength

Oleg Popov[1] ✉, ov.popov@mail.ru
Andrey Tumashov[1], ice47reg@yandex.ru
Georgy Borisov[1], georgiiborisov@gmail.com
Konstantin Korovin[2], korovin.ko@sut.ru

[1]Special Technology Center LLC,
St. Petersburg, 195220, Russian Federation
[2]The Bonch-Bruevich Saint-Petersburg State University of Telecommunications,
St. Petersburg, 193232, Russian Federation

Abstract: *A method for calculating the complex capacitance of small with respect to the wavelength, normally open antennas is proposed. Analytical expressions for determination of mutual potential coefficient of any two arbitrarily arranged linear conductors are obtained. The proposed technique can be useful in assessing the losses of normally open wire antennas placed in close proximity to the semiconductor surface, as well as in determination of their input resistance.*

Keywords: *complex capacitance, mutual potential coefficient, unbalanced monopole*

For citation: Popov O., Tumashov A., Borisov G., Korovin K. Mathematical Model of the Unbalanced Monopole Feed. Part 2. Determination of Complex Capacitance of Normally Open Wire Antennas, Small with Respect to the Wavelength. *Proc. of Telecom. Universities.* 2023;9(2):6–21. (in Russ.) DOI:10.31854/1813-324X-2023-9-2-6-21

Введение

Комплексной емкостью (КЕ), как правило, характеризуются нормально разомкнутые антенны, максимальный габарит которых существенно меньше длины волны. Обычно это развертываемые в непосредственной близости от границы раздела с полупроводящей средой, антенны ВЧ-диапазона и более низких частот. Антенна при этом может находиться как в воздухе, так и в полупроводящей среде (земле, воде). Возможен и промежуточный случай, когда часть конструкции антенны находится в одной среде и часть – в другой. При любом варианте размещения, ближнее поле антенны наводит в полупроводящей среде токи проводимости, преобразующие энергию электромагнитного поля в тепловую. Зона возникающих тепловых потерь охватывает расстояния, равные примерно сумме горизонтального размера антенны и ее высоты. За пределами этой зоны поле уже не связано с антенной и возникающие здесь потери относятся к потерям в тракте распространения. Таким образом, КЕ описывает как реактивную составляющую входного сопротивления, так и сопротивление потерь в подстилающей поверхности и элементах конструкции, что делает задачу определения КЕ весьма полезной как при проектировании антенн, так и при оценке характеристик радиокомплексов, в которых они применяются [1].

Как известно [2–4], нормально разомкнутая антенна представляет собой совокупность двух не имеющих между собой контакта проводящих тел, определенным образом подобранных и взаимно ориентированных. Обычно эти тела называют плечами антенны.

КЕ, согласно определению [2–4], называют заряд, накапливаемый на каждом плече антенны при единичной разности потенциалов между плечами. Из электростатики известно [5], что потенциал всех точек тела, имеющего не нулевую проводимость, одинаков. Следовательно, заряды по поверхности каждого плеча антенны должны быть распределены таким образом, чтобы выполнялось условие:

$$\varphi_p = \frac{1}{4\pi\varepsilon_a}\int\limits_S \frac{\delta_q}{r_{pq}} dS_q = \text{const}, \qquad (1)$$

где φ_p – потенциал в точке p; δ_q – поверхностная плотность заряда в точке q; r_{pq} – расстояние между произвольными точками p и q, лежащими на поверхности плеча антенны; ε_a – диэлектрическая проницаемость окружающей среды; S – поверхность плеча антенны.

Решение интегрального уравнения (1) представляет собой сложную математическую задачу, в которой обычно используются приближенные методы. Наибольшее распространение получил метод среднего потенциала или метод Хоу [6, 7]. Этот метод основан на том, что плотность заряда на поверхности длинного проводника практически постоянна, за исключением его концов, где она резко возрастает. На этом основании считают, что поверхностный заряд по всему проводнику распределяют равномерно, а его потенциал соответствует среднему значению потенциала во всех точках поверхности проводника.

В этом случае потенциал точки p, как следует из (1), будет следующим:

$$\varphi_p = \frac{\delta_q}{4\pi\varepsilon_a}\int\limits_S \frac{dS_q}{r_{pq}}. \qquad (2)$$

Среднее значение потенциала оказывается весьма близким к истинному потенциалу проводника на большей части его поверхности:

$$\overline{\varphi} = \frac{1}{S}\int\limits_S \varphi_p dS_p = \frac{Q}{4\pi\varepsilon_a S^2}\int\limits_S\int\limits_S \frac{dS_q dS_p}{r_{pq}}. \qquad (3)$$

Емкость одиночного проводника в этом случае определится как отношение заряда к среднему потенциалу:

$$C = \frac{Q}{\overline{\varphi}} = \frac{4\pi\varepsilon_a S^2}{\int_S\int_S \frac{dS_q dS_p}{r_{pq}}}. \qquad (4)$$

Однако методом среднего потенциала, в измененном виде, можно обеспечить достаточно точный результат расчета емкости плеча формально разомкнутой антенны только в том случае, когда распределение заряда по его поверхности близко к равномерному. Если плечо состоит из нескольких,

отличающихся друг от друга частей, то расчет емкости в предположении равномерного распределения заряда приводит к значительным ошибкам. В этом случае плотность распределения заряда следует считать равномерной не для всего плеча, а для каждой из его частей [6]. Среднее же значение потенциала каждой части антенны должно определяться собственным зарядом этой части, а также зарядами всех остальных частей при удовлетворении граничным условиям на поверхности полупроводящей земли.

Выполнение граничных условий обеспечивается обобщением метода зеркальных изображений [2, 5, 8] на квазистатическую область. В электростатике такой метод используется для расчета поля стационарных зарядов при наличии плоской границы раздела между двумя средами. Для нахождения поля переменных зарядов метод зеркальных изображений, в строгой постановке, применим лишь в случае границы с идеально проводящей средой. Однако, если ограничиться определением поля в области с максимальным габаритом не более четверти длины волны, то данный метод можно обобщить и на полупроводящие среды [8, 9].

Комплексная емкость нормально разомкнутой проволочной антенны

Составными частями проволочных антенн удобно считать образующие их проводники. Следовательно, применительно к проволочным антеннам суть метода состоит в том, что для выполнения граничных условий на поверхности земли достаточно дополнить систему проводников их зеркальными изображениями, относительно границы раздела сред, а также ввести в рассмотрение понятие «виртуальный заряд». При этом среду, в которой размещается проводник, потенциал которого определяется, называют средой оригинала, а среду с изображением этого проводника – средой изображения.

Поле в среде оригинала представляется в виде суперпозиции полей всех проводников, находящихся в этой среде, и их изображений, при условии, что, как проводники, так и их изображения находятся в однородной среде с параметрами среды оригинала. Заряды проводников изображений связаны с зарядами оригиналов соотношениями [2, 10]:

$$Q_{\underline{N}}^{\text{и}} = \gamma Q_{N}^{\text{о}}, \tag{5}$$

$$Q_{N}^{\text{и}} = -\gamma Q_{\underline{N}}^{\text{о}}, \tag{6}$$

где

$$\gamma = \frac{1-\varepsilon_k}{1+\varepsilon_k};\ \varepsilon_k = \varepsilon - j60\sigma\lambda; \tag{7}$$

$Q_{\underline{N}}^{\text{и}}$ – заряд, находящегося в земле изображения N-го проводника; $Q_{N}^{\text{о}}$ – заряд N-го проводника, находящегося в воздухе (заряд оригинала); $Q_{N}^{\text{и}}$ – заряд, находящегося в воздухе изображения N-го проводника; $Q_{\underline{N}}^{\text{о}}$ – заряд N-го проводника, находящегося в земле (заряд оригинала); γ – коэффициент отражения статического заряда; ε_k – относительная комплексная диэлектрическая проницаемость земли; ε – относительная диэлектрическая проницаемость земли; σ – удельная проводимость земли; λ – длина волны.

Верхний индекс заряда обозначает его природу (о – оригинал; и – изображение). Нижний индекс обозначает номер проводника, по поверхности которого этот заряд распределен. Знак тильда над нижним индексом означает, что данный проводник находится над границей раздела, а под нижним индексом – под границей раздела.

Поле в среде изображения, согласно методу зеркальных изображений, представляется в виде суперпозиции полей, созданных некими виртуальными зарядами, распространенными по поверхности проводников, находящимся в одной среде с проводником оригиналом. При этом считается, что все эти проводники расположены в бесконечной среде с параметрами среды изображения.

Виртуальные заряды связаны с зарядами проводников оригиналов соотношениями [2, 10]:

$$Q_{N}^{\text{в}} = (1-\gamma)Q_{N}^{\text{о}}, \tag{8}$$

$$Q_{\underline{N}}^{\text{в}} = (1+\gamma)Q_{\underline{N}}^{\text{о}}, \tag{9}$$

где $Q_{N}^{\text{в}}$ – виртуальный заряд N-го проводника, находящегося в воздухе; $Q_{\underline{N}}^{\text{в}}$ – виртуальный заряд N-го проводника, находящегося в земле; верхний индекс означает природу заряда (в – виртуальный).

Соотношения (5–9) позволяют установить связь между потенциалами проводников проволочной антенны и находящимися на них зарядами. Рассмотрим некоторую совокупность проводников, не имеющих между собой электрического контакта. Пусть N проводников находится над поверхностью земли, а M проводников – под поверхностью. Тогда потенциал $\bar{\imath}$-го изолированного проводника, расположенного над поверхностью земли, можно представить тремя суммами потенциалов [2]:

$$\varphi_{\bar{\imath}} = \sum_{n=1}^{N} \varphi_{\bar{\imath}\,\bar{n}} + \sum_{n=1}^{N} \varphi_{\bar{\imath}\,\underline{n}} + \sum_{m=1}^{M} \varphi_{\bar{\imath}\,\underline{m}}, \tag{10}$$

где $\varphi_{\bar{\imath}}$ – потенциал $\bar{\imath}$-го проводника, находящегося над поверхностью земли; $\varphi_{\bar{\imath}\,\bar{n}}$ – потенциал, созданный на $\bar{\imath}$-м проводнике зарядом $Q_{\bar{n}}^{\text{о}}$, находящимся на $\bar{n}$-м проводнике, расположенном над поверхностью земли; $\varphi_{\bar{\imath}\,\underline{n}}$ – потенциал, созданный на $\bar{\imath}$-м проводнике, зарядом изображения $\underline{n}$-го проводника $Q_{\underline{n}}^{\text{и}}$ (изображение находится под поверхностью земли, на что указывает знак тильда); $\varphi_{\bar{\imath}\,\underline{m}}$ – потенциал, созданный на $\bar{\imath}$-м проводнике виртуальным зарядом $Q_{\underline{m}}^{\text{в}}$, находящимся на проводнике $\underline{m}$, расположенным под поверхностью земли.

Поскольку потенциалы $\varphi_{\bar{i}\,\bar{n}}$, $\varphi_{\bar{i}\,\underline{n}}$ и $\varphi_{\bar{i}\,\underline{m}}$ пропорциональны соответствующим зарядам $Q^{\text{о}}_{\bar{n}}$, $Q^{\text{и}}_{\underline{n}}$, и $Q^{\text{в}}_{\underline{m}}$, соотношение (10) можно представить в виде:

$$\varphi_{\bar{i}} = \sum_{n=1}^{N} p_{\bar{i}\,\bar{n}} Q^{\text{о}}_{\bar{n}} + \sum_{n=1}^{N} p_{\bar{i}\,\underline{n}} Q^{\text{и}}_{\underline{n}} + \sum_{m=1}^{M} p_{\bar{i}\,\underline{m}} Q^{\text{в}}_{\underline{m}}, \qquad (11)$$

где $p_{\bar{i}\,\bar{n}}$, $p_{\bar{i}\,\underline{n}}$, и $p_{\bar{i}\,\underline{m}}$ – потенциалы, наведенные на проводнике $\bar{i}$, единичными зарядами, расположенными на проводниках $\bar{n}$, $\underline{n}$ или $\underline{m}$, при условии, что оба проводника находятся в однородной безграничной среде, параметры которой совпадают с параметрами среды проводника $\bar{i}$ (среды, на которую указывает знак тильды первого нижнего индекса). Эти потенциалы называются потенциальными коэффициентами. Коэффициенты, у которых первый нижний индекс полностью совпадает со вторым (например, $p_{\bar{i}\,\bar{i}}$) называется собственными, а все остальные – взаимными.

Легко показать, что если проводник $\underline{j}$ находится под поверхностью земли, то соотношение, описывающее связь его потенциала с зарядами всех проводников, будет следующим:

$$\varphi_{\underline{j}} = \sum_{m=1}^{M} p_{\underline{j}\,\underline{m}} Q^{\text{о}}_{\underline{m}} + \sum_{m=1}^{M} p_{\underline{j}\,\bar{m}} Q^{\text{и}}_{\bar{m}} + \sum_{n=1}^{N} p_{\underline{j}\,\bar{n}} Q^{\text{в}}_{\bar{n}}. \qquad (12)$$

Видно, что все потенциальные коэффициенты, входящие в соотношение (12), определяются для проводников, находящихся в безграничной однородной среде с параметрами земли.

Соотношения (5, 6 и 8, 9) позволяют перейти в (11 и 12) от зарядов изображений и виртуальных зарядов к зарядам проводников оригиналов. Получим:

$$\varphi_{\bar{i}} = \sum_{n=1}^{N} \left(p_{\bar{i}\,\bar{n}} + \gamma p_{\bar{i}\,\underline{n}}\right) Q^{\text{о}}_{\bar{n}} + \sum_{m=1}^{M} (1+\gamma) p_{\bar{i}\,\underline{m}} Q^{\text{о}}_{\underline{m}}, \qquad (13)$$

$$\varphi_{\underline{j}} = \sum_{n=1}^{N} (1-\gamma) p_{\underline{j}\,\bar{n}} Q^{\text{о}}_{\bar{n}} + \sum_{m=1}^{M} \left(p_{\underline{j}\,\underline{m}} - \gamma\right) p_{\underline{j}\,\bar{m}} Q^{\text{о}}_{\underline{m}}. \qquad (14)$$

Полученные выражения для дальнейшего анализа удобно представить в более компактном виде:

$$\varphi_{\bar{i}} = \sum_{n=1}^{N} p^{\Sigma}_{\bar{i}\,\bar{n}} Q^{\text{о}}_{\bar{n}} + \sum_{m=1}^{M} p^{\Sigma}_{\bar{i}\,\underline{m}} Q^{\text{о}}_{\underline{m}}, \qquad (15)$$

$$\varphi_{\underline{j}} = \sum_{n=1}^{N} p^{\Sigma}_{\underline{j}\,\bar{n}} Q^{\text{о}}_{\bar{n}} + \sum_{m=1}^{M} p^{\Sigma}_{\underline{j}\,\underline{m}} Q^{\text{о}}_{\underline{m}}, \qquad (16)$$

где

$$p^{\Sigma}_{\bar{i}\,\bar{n}} = p_{\bar{i}\,\bar{n}} + \gamma p_{\bar{i}\,\underline{n}}; \qquad (17)$$

$$p^{\Sigma}_{\bar{i}\,\underline{m}} = (1+\gamma) p_{\bar{i}\,\underline{m}}; \qquad (18)$$

$$p^{\Sigma}_{\underline{j}\,\bar{n}} = (1-\gamma) p_{\underline{j}\,\bar{n}}; \qquad (19)$$

$$p^{\Sigma}_{\underline{j}\,\underline{m}} = p_{\underline{j}\,\underline{m}} - \gamma p_{\underline{j}\,\bar{m}}; \qquad (20)$$

$p^{\Sigma}_{\bar{i}\,\bar{n}}$, $p^{\Sigma}_{\bar{i}\,\underline{m}}$ – потенциальные коэффициенты, представляющие собой потенциал проводника $\bar{i}$, находящегося над полупроводящей землей, созданный единичным зарядом, распределенным по проводнику $\bar{n}$, находящемуся над поверхностью или по проводнику $\underline{m}$ – под поверхностью земли, соответственно; $p^{\Sigma}_{\underline{j}\,\bar{n}}$, $p^{\Sigma}_{\underline{j}\,\underline{m}}$ – потенциальные коэффициенты, представляющие собой потенциал проводника $\underline{j}$, находящегося в полупроводящей земле, созданный единичным зарядом, распределенным по проводнику $\bar{n}$, находящемуся над поверхностью или по проводнику $\underline{m}$ – под поверхностью земли, соответственно.

Поскольку соотношения (15) и (16) могут использоваться для определения потенциалов любого проводника, расположенного над поверхностью раздела сред и под поверхностью, соответственно, их удобно представить в виде матричной системы уравнений:

$$\left.\begin{aligned} \varphi_{\bar{N}}\rangle &= \left|P^{\Sigma}_{\bar{N}\,\bar{N}}\right| Q^{\text{о}}_{\bar{N}}\rangle + \left|P^{\Sigma}_{\bar{N}\,\underline{M}}\right| Q^{\text{о}}_{\underline{M}}\rangle \\ \varphi_{\underline{M}}\rangle &= \left|P^{\Sigma}_{\underline{M}\,\bar{N}}\right| Q^{\text{о}}_{\bar{N}}\rangle + \left|P^{\Sigma}_{\underline{M}\,\underline{M}}\right| Q^{\text{о}}_{\underline{M}}\rangle \end{aligned}\right\}, \qquad (21)$$

где $\varphi_{\bar{N}}\rangle$ – матрица-столбец потенциалов N проводников, находящихся над поверхностью земли; $\varphi_{\underline{M}}\rangle$ – матрица-столбец потенциалов M проводников, находящихся под поверхностью земли; $Q^{\text{о}}_{\bar{N}}\rangle$ – матрица-столбец, распределенных по N проводникам, находящимся над поверхностью земли; $Q^{\text{о}}_{\underline{M}}\rangle$ – матрица-столбец зарядов, распределенных по M проводникам, находящимся под поверхностью земли; $\left|P^{\Sigma}_{\bar{N}\,\bar{N}}\right|$ – квадратная матрица $N \times N$ взаимных потенциальных коэффициентов проводников, находящихся над поверхностью земли; $\left|P^{\Sigma}_{\bar{N}\,\underline{M}}\right|$ – прямоугольная матрица $N \times M$ взаимных потенциальных коэффициентов, устанавливающая связь потенциалов проводников расположенными над поверхностью земли с зарядами на проводниках, находящихся в земле; $\left|P^{\Sigma}_{\underline{M}\,\bar{N}}\right|$ – прямоугольная матрица $M \times N$ взаимных потенциальных коэффициентов, устанавливающих связь потенциалов проводников, находящихся в земле, с зарядами на проводниках, расположенных над ее поверхностью; $\left|P^{\Sigma}_{\underline{M}\,\underline{M}}\right|$ – квадратная матрица $M \times M$ взаимных потенциальных коэффициентов проводников под поверхностью земли.

Систему (21) можно представить в виде единого матричного уравнения:

$$\left|\begin{matrix}\varphi_{\bar{N}}\rangle \\ \varphi_{\underline{M}}\rangle\end{matrix}\right| = \left|\begin{matrix}\left|P^{\Sigma}_{\bar{N}\,\bar{N}}\right| & \left|P^{\Sigma}_{\bar{N}\,\underline{M}}\right| \\ \left|P^{\Sigma}_{\underline{M}\,\bar{N}}\right| & \left|P^{\Sigma}_{\underline{M}\,\underline{M}}\right|\end{matrix}\right| \left|\begin{matrix}Q^{\text{о}}_{\bar{N}}\rangle \\ Q^{\text{о}}_{\underline{M}}\rangle\end{matrix}\right| \qquad (22)$$

или

$$\varphi_{K}\rangle = \left|P^{\Sigma}_{K\,K}\right| Q^{\text{о}}_{K}\rangle,$$

где $K = N + M$ – общее число проводников в проволочной антенне; $\left|P_{K\ K}^{\Sigma}\right|$ – обобщенная квадратная матрица взаимных потенциальных коэффициентов проводников (МВПК) проволочной антенны.

Как видно из (22), при условии изолированности всех проводников, распределенные по ним заряды связаны с потенциалами соотношением:

$$Q_K\rangle = |C_{K\ K}|\varphi_K\rangle, \qquad (23)$$

где $|C_{K\ K}| = \left|P_{K\ K}^{\Sigma}\right|^{-1}$ – квадратная матрица $K \times K$ взаимных емкостей проводников проволочной антенны.

Взаимной емкостью C_{ij} называют заряд, накапливаемый на проводнике i под воздействием единичного потенциала проводника j. Емкость C_{ii} называют собственной емкостью проводника i.

Для определения КЕ на соотношение (23) необходимо наложить граничные условия. Прежде всего, следует учесть, что все проводники нормально разомкнутой антенны объединены в два плеча, причем проводники, образующие одно плечо, имеют между собой электрический контакт и, следовательно, равные потенциалы.

Пусть первое и второе плечи антенны содержат V и T проводников, соответственно, причем $V + T = K$. Тогда соотношение (23) можно представить в виде системы матричных уравнений:

$$\left.\begin{aligned} Q_V\rangle &= |C_{V\ V}|1_V\rangle\varphi_V + |C_{V\ T}|1_T\rangle\varphi_T \\ Q_T\rangle &= |C_{T\ V}|1_V\rangle\varphi_V + |C_{T\ T}|1_T\rangle\varphi_T \end{aligned}\right\}, \qquad (24)$$

где $Q_V\rangle, Q_T\rangle$ – матрица-столбец зарядов проводников первого и второго плеча, соответственно; $1_V\rangle$, $1_T\rangle$ – единичная матрица-столбец из V и T элементов, соответственно; $|C_{VV}|$ – квадратная матрица $V \times V$ собственных и взаимных емкостей проводников первого плеча; $|C_{VT}|$ – прямоугольного матрица $V \times T$ взаимных емкостей проводников первого и второго плечей; $|C_{TV}|$ – прямоугольная матрица $T \times V$ взаимных емкостей проводников второго и первого плечей; $|C_{TT}|$ – квадратная матрица $T \times T$ собственных и взаимных емкостей проводников второго плеча; φ_V, φ_T – потенциалы первого и второго плечей антенны, соответственно.

Умножая квадратные и прямоугольные матрицы на единичные матрицы столбцы в правой части соотношения (24), легко найти заряды отдельных проводников первого и второго плечей нормально разомкнутой проволочной антенны:

$$\left.\begin{aligned} {Q_i}^V\rangle &= \varphi_V \sum_{v=1}^{V} C_{i\,v} + \varphi_T \sum_{t=1}^{T} C_{i\,t} \\ {Q_j}^T\rangle &= \varphi_V \sum_{v=1}^{V} C_{jv} + \varphi_T \sum_{t=1}^{T} C_{j\,t} \end{aligned}\right\}, \qquad (25)$$

где ${Q_i}^V$ – заряд i-го проводника первого плеча антенны; ${Q_j}^T$ – заряд j-го проводника второго плеча антенны.

Суммарные заряды всех проводников первого и второго плечей антенны будут:

$$\left.\begin{aligned} {Q_\Sigma}^V\rangle &= \varphi_V \sum_{i=1}^{V}\sum_{v=1}^{V} C_{i\,v} + \varphi_T \sum_{i=1}^{V}\sum_{t=1}^{T} C_{i\,t} \\ {Q_\Sigma}^T\rangle &= \varphi_V \sum_{j=1}^{T}\sum_{v=1}^{V} C_{jv} + \varphi_T \sum_{j=1}^{T}\sum_{t=1}^{T} C_{j\,t} \end{aligned}\right\}. \qquad (26)$$

Поскольку заряды, накопленные на плечах нормально разомкнутой антенны, должны быть равны по величине и противоположны по знаку, следовательно:

$${Q_\Sigma}^V = -{Q_\Sigma}^T = Q. \qquad (27)$$

Решая систему (26), с учетом соотношения (27), относительно зарядов плечей антенны, получим:

$$\varphi_T = Q\frac{C_{VV} + C_{VT}}{C_{TV}C_{VT} - C_{TT}C_{VV}}, \qquad (28)$$

$$\varphi_V = -Q\frac{C_{TT} + C_{VT}}{C_{TV}C_{VT} - C_{TT}C_{VV}}, \qquad (29)$$

где

$$C_{VV} = \sum_{i=1}^{V}\sum_{v=1}^{V} C_{i\,v}, \qquad (30)$$

$$C_{VT} = \sum_{i=1}^{V}\sum_{t=1}^{T} C_{i\,t}, \qquad (31)$$

$$C_{TV} = \sum_{j=1}^{T}\sum_{v=1}^{V} C_{jv}, \qquad (32)$$

$$C_{TT} = \sum_{j=1}^{T}\sum_{t=1}^{T} C_{jt}. \qquad (33)$$

Соотношения (28 и 29) позволяют определить в общем виде КЕ нормально разомкнутой проволочной антенны, габариты которой много меньше длины волны.

По определению, КЕ есть отношение заряда на одном из плечей антенны к разности потенциалов между плечами:

$$C = \frac{Q}{\varphi_V - \varphi_T} = \frac{C_{TT}C_{VV} - C_{TV}C_{VT}}{C_{TT} + C_{VT} + C_{TV} + C_{VV}}. \qquad (34)$$

Таким образом, для определения КЕ нормально разомкнутой антенны необходимо виртуально расчленить ее конструкцию на отдельные фрагменты, в пределах которых распределение заряда по поверхности можно считать постоянным. Затем, считая, что электрический контакт между фрагментами отсутствует, определить собственные и взаимные потенциальные коэффициенты, как между

самими фрагментами, так и между фрагментами и их изображениями. Эта задача должна решаться при условии, что каждая пара фрагмент – изображение находится в однородной бесконечной среде с параметрами среды оригинала. Для каждой пары фрагментов, находящихся в разных средах, следует определить два взаимных потенциальных коэффициента – один для случая, когда оба фрагмента находятся в земле, другой – в воздухе. После этого последовательное применение соотношений (17–23), а затем (30–34) позволяет вычислить КЕ нормально разомкнутой антенны выбранной конструкции.

Исходя из изложенного, одной из основных задач, требующих решения при вычислении КЕ проволочной антенны произвольной формы, является определение взаимного потенциального коэффициента двух произвольно ориентированных проводников. Такая задача рассматривалась в литературе. Однако приведенные в [3 и 14] аналитические выражения даны без вывода и не сходятся между собой, что говорит о наличии опечаток как минимум в одном из источников. Учитывая важность получения достоверного результата, представляется полезным повторное решение этой задачи и подробное его описание, допускающее независимую проверку.

Взаимный потенциальный коэффициент произвольно ориентированных проводников

В качестве предварительного замечания следует отметить, что через оси двух произвольно ориентированных проводников, продолжения которых не пересекаются, всегда можно провести две параллельные плоскости. Действительно, выберем произвольную точку на оси одного проводника и проведем через нее вспомогательную прямую параллельно оси второго проводника. Плоскость, задаваемая этой прямой и осью первого проводника, будет параллельна оси второго проводника. Аналогичным образом выберем точку на оси второго проводника и проведем через нее вспомогательную прямую параллельно оси первого проводника. Легко доказать, что плоскость, в которой лежит ось второго проводника и пересекающая ее вторая вспомогательная линия, параллельна плоскости, задаваемой осью первого проводника и первой вспомогательной линией.

На рисунке 1 изображены оси двух проводников, расположенных в двух параллельных плоскостях, находящихся на расстоянии d друг от друга. Выберем систему координат таким образом, чтобы ось x совпадала с осью первого проводника. Путем параллельного переноса вдоль оси y создадим в плоскости второго проводника вспомогательную систему координат $x'o'z'$. Тогда угол между осью второго проводника и осью x' будет равен α.

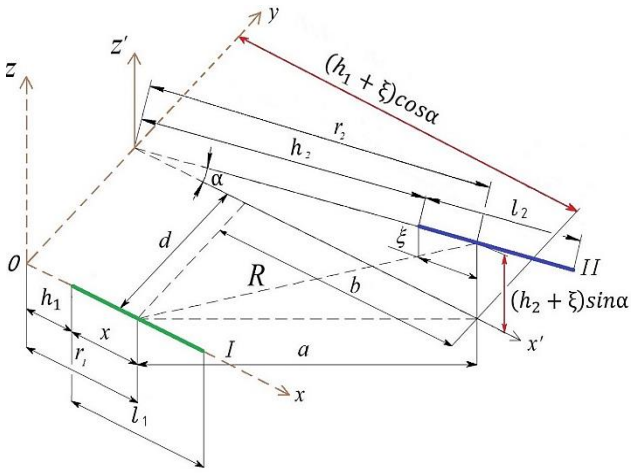


Рис. 1. Оси произвольно расположенных проводников в двух параллельных плоскостях

Fig.1. The Axis of Arbitrary Placed Conductors in Two Parallel Planes

Пусть h_1 и h_2 – расстояния от ближайших к оси y торцов первого и второго проводников, соответственно, до точек пересечения продолжений осей этих проводников с осью y. Выберем на оси первого проводника точку, отстоящую на удалении x, а на оси второго проводника – на удалении ξ от торцов, ближайших к оси y.

Из рисунка 1 видно, что расстояние между этими точками будет:

$$R = \sqrt{a^2 + {r_2}^2 \sin^2\alpha}, \qquad (35)$$

где $a = \sqrt{b^2 + d^2}$; $b = r_2\cos\alpha - r_1$; $r_1 = h_1 + x$; $r_2 = h_2 + \xi$.

Выражение (35) с учетом сделанных обозначений удобно представить в виде функции двух переменных r_1 и r_2, характеризующих расстояния до оси y от точек x и ξ, соответственно:

$$R(r_1, r_2; d, \alpha) = \sqrt{{r_1}^2 + {r_2}^2 + d^2 - 2r_1 r_2\cos\alpha}. \qquad (36)$$

В тех случаях, когда радиусы проводников много меньше их длин, можно, практически без потери точности, определять средние потенциалы путем усреднения их значений не по поверхностям, а по осям проводников [2]. Будем считать, что заряд Q_2 равномерно распределен по оси второго проводника. В этом случае можно считать, что в точке ξ находится элементарный заряд $(Q_2/l_2)d\xi$, который создает на оси первого проводника в точке x потенциал [12]:

$$d\varphi_{12}(x) = \frac{Q_2 d\xi}{4\pi\varepsilon_a l_2 R(r_1, r_2, d, \alpha)}, \qquad (37)$$

где $\varepsilon_a = \varepsilon_0\varepsilon_k$ – абсолютная диэлектрическая проницаемость среды, в которой находится проводник; $\varepsilon_0 = \frac{1}{36\pi}10^{-9}$ [Ф/м] – диэлектрическая проницаемость вакуума.

Вся совокупность элементарных зарядов, равномерно распределенных по второму проводнику, создает в точке x потенциал, равный:

$$\varphi_{12}(x) = \frac{Q_2}{4\pi\varepsilon_a l_2}\int_0^{l_2}\frac{d\xi}{R(r_1, h_2+\xi, d, \alpha)}. \quad (38)$$

Для вычисления интеграла (38) введем в рассмотрение новую переменную:

$$\eta = h_2 + \xi - r_1\cos\alpha + R(r_1, h_2+\xi, d, \alpha).$$

Тогда с учетом (35–36), дифференциал этой переменной будет следующим:

$$d\eta = \left[1 + \frac{h_2+\xi-r_1\cos\alpha}{R(r_1, h_2+\xi, d, \alpha)}\right]d\xi = \frac{\eta d\xi}{R(r_1, h_2+\xi, d, \alpha)}.$$

Отсюда следует:

$$\frac{d\xi}{R(r_1, h_2+\xi, d, \alpha)} = \frac{d\eta}{\eta}. \quad (39)$$

Выражение (39) позволяет легко найти аналитическое представление интеграла (38):

$$\varphi_{12}(x) = \frac{Q_2}{4\pi\varepsilon_a l_2}\int_{\eta(0)}^{\eta(l_2)}\frac{d\eta}{\eta} = \frac{Q_2}{4\pi\varepsilon_a l_2}\ln\frac{\eta(l_2)}{\eta(0)}, \quad (40)$$

где

$$\eta(l_2) = h_2 + l_2 - r_1\cos\alpha + R(r_1, h_2+\xi, d, \alpha);$$
$$\eta(0) = h_2 - r_1\cos\alpha + R(r_1, h_2+\xi, d, \alpha).$$

Средний потенциал первого проводника, созданный равномерным распределением заряда Q_2 по второму проводнику, будет вычисляться по выражению (41). Соотношение (41) можно записать в более компактном виде (42), сделав замену переменных $x = r_1 - h$.

$$\varphi_{12}(x) = \frac{Q_2}{4\pi\varepsilon_a l_1 l_2}\left\{\int_0^{l_1}\ln|h_2 + l_2 - (h_1+x)\cos\alpha + R(h_1+x, h_2+l_2, d, \alpha)|dx - \right.$$
$$\left. - \int_0^{l_1}\ln|h_2 - (h_1+x)\cos\alpha + R(h_1+x, h_2, d, \alpha)|dx\right\}. \quad (41)$$

$$\varphi_{12}(x) = \frac{Q_2}{4\pi\varepsilon_a l_1 l_2}\left\{\int_{h_1}^{h_1+l_1}\ln|h_2 + l_2 - r_1\cos\alpha + R(r_1, h_2+l_2, d, \alpha)|dr_1 - \right.$$
$$\left. - \int_{h_1}^{h_1+l_1}\ln|h_2 - r_1\cos\alpha + R(r_1, h_2, d, \alpha)|dr_1\right\}. \quad (42)$$

Легко заметить, что вычисление интегралов в выражении (42) сводится к вычислению неопределенного интеграла следующего вида:

$$J(y; a, d, \alpha) = \int \ln|a - y\cos\alpha + R(y, a; d, \alpha)|dy, \quad (43)$$

где

$$R(y, a; d, \alpha) = \sqrt{y^2 + a^2 + d^2 - 2ay\cos\alpha}. \quad (44)$$

Вычисление интеграла (43) можно начать с применения к нему правила интегрирования по частям:

$$J(y; a, d, \alpha) = A(y; a, d, \alpha) - I1(y; a, d, \alpha), \quad (45)$$
$$A(y; a, d, \alpha) = y\ln|a - y\cos\alpha + R(y; a, d, \alpha)|, \quad (46)$$
$$I1(y; a, d, \alpha) = \int \frac{-\cos\alpha + \frac{d}{dy}R(y, a; d, \alpha)}{a - y\cos\alpha + R(y, a; d, \alpha)}ydy. \quad (47)$$

Поскольку:

$$\frac{d}{dy}R(y, a; d, \alpha) = \frac{y - a\cos\alpha}{R(y, a; d, \alpha)},$$

интеграл (47) будет иметь вид:

$$I1(y; a, d, \alpha) =$$
$$= \int \frac{y^2 - ay\cos\alpha - yR(y, a; d, \alpha)\cos\alpha}{R(y, a; d, \alpha)[a - y\cos\alpha + R(y, a; d, \alpha)]}dy. \quad (48)$$

Подынтегральное выражение в (48) можно существенно упростить, добавив к числителю и одновременно вычтя из него следующую сумму: $a^2 + d^2 + ay\cos\alpha - aR(y, a; d, \alpha)$. Тогда, учитывая обозначение (44), получим (49). Легко заметить, что интеграл (48) с учетом (49) распадается на сумму трех более простых интегралов (50).

$$y^2 - ay\cos\alpha - yR(y, a; d, \alpha)\cos\alpha = y^2 + a^2 + d^2 - 2ay\cos\alpha - yR(y, a; d, \alpha)\cos\alpha + aR(y, a; d, \alpha) -$$
$$-a^2 + ay\cos\alpha - aR(y; a, d, \alpha) - d^2 = R(y, a; d, \alpha)[a - y\cos\alpha + R(y, a; d, \alpha)] -$$
$$-a[a - y\cos\alpha + R(y, a; d, \alpha)] - d^2. \quad (49)$$

$$I1(y;a,d,\alpha)=\int dy-a\int\frac{dy}{R(y,a;d,\alpha)}-d^2\int\frac{dy}{R(y,a;d,\alpha)[a-y\cos\alpha+R(y,a;d,\alpha)]}= \quad (50)$$

$$=y-a\ I11(y;a,d,\alpha)-d^2\ I12(y;a,d,\alpha),$$

где

$$I11(y;a,d,\alpha)=\int\frac{dy}{R(y,a;d,\alpha)}; \quad (51)$$

$$I12(y;a,d,\alpha)=\int\frac{dy}{R(y,a;d,\alpha)[a-y\cos\alpha+R(y,a;d,\alpha)]}. \quad (52)$$

Интеграл (51) можно вычислить, сделав замену переменной двумя способами:

$$y=a\cos\alpha-z \quad (53)$$

и

$$y=z-a\cos\alpha. \quad (54)$$

При замене (53) интеграл (51) с учетом (44) приобретает вид выражения (55). Можно видеть, что интеграл (55) является частным случаем (см. выражение 56) известного интеграла [13, Инт.1.2.52.8].

С учетом соотношения (56), аналитическое значение интеграла (55) будет представлено в следующем виде:

$$I11(z';a,d,\alpha)=-\ln\left|z'+\sqrt{(z')^2+a^2\sin^2\alpha+d^2}\right|.$$

Делая обратную замену переменных, интеграл будет иметь вид (57). Замена переменных (54) преобразует интеграл (51) к несколько иному виду (58). Приводя соотношение (58) аналогичным образом к аналитическому виду и делая обратную замену переменной, получим выражение (59). Несмотря на видимое различие соотношений (57 и 59), легко показать, что они совпадают с точностью до константы. Действительно, преобразуем соотношение (57) к следующему виду (60). Таким образом, цепочка преобразований (60) показывает, что при вычислении определенных интегралов в (42) может использоваться как представление (57), так и (59), поскольку оба они дадут одинаковый конечный результат.

$$I11(z';a,d,\alpha)=-\int\frac{dz'}{\sqrt{a^2\cos\alpha-2az'\cos\alpha+(z')^2+a^2+d^2-2a(a\cos\alpha-z')\cos\alpha}}= \quad (55)$$

$$=-\int\frac{dz'}{\sqrt{(z')^2+a^2\sin^2\alpha+d^2}}.$$

$$\int\frac{dx}{\sqrt{ax^2+bx+c}}=\frac{1}{\sqrt{a}}\ln\left|\frac{2ax+b}{2\sqrt{a}}+\sqrt{ax^2+bx+c}\right|. \quad (56)$$

$$I11(z';a,d,\alpha)=-\ln\left|a\cos\alpha-y+\sqrt{a^2+d^2+y^2-2ay\cos\alpha}\right|=-\ln|a\cos\alpha-y+R(y;a,d,\alpha)|. \quad (57)$$

$$\bar{I}11(z";a,d,\alpha)=-\int\frac{dz"}{\sqrt{a^2\cos\alpha-2az"\cos\alpha+(z")^2+a^2+d^2-2a(a\cos\alpha-z")\cos\alpha}}= \quad (58)$$

$$=-\int\frac{dz"}{\sqrt{(z")^2+a^2\sin^2\alpha+d^2}}.$$

$$\bar{I}11(y;a,d,\alpha)=\ln\left|y-a\cos\alpha\sqrt{a^2+d^2+y^2-2ay\cos\alpha}\right|=\ln|y-a\cos\alpha+R(y,a;d,\alpha)|. \quad (59)$$

$$\bar{I}11(y;a,d,\alpha)=-\ln\left|\frac{[a\cos\alpha-y+R(y,a;d,\alpha)][a\cos\alpha-y-R(y,a;d,\alpha)]}{a\cos\alpha-y-R(y,a;d,\alpha)}\right|=$$

$$=-\ln\left|\frac{a^2\cos^2\alpha-2ay\cos\alpha+y^2-a^2-d^2-y^2+2ay\cos\alpha}{a\cos\alpha-y-R(y,a;d,\alpha)}\right|= \quad (60)$$

$$=-\ln\left|\frac{-a^2\sin^2\alpha-d^2}{a\cos\alpha-y-R(y,a;d,\alpha)}\right|=-\ln\left|\frac{a^2\sin^2\alpha+d^2}{y-a\cos\alpha+R(y,a;d,\alpha)}\right|=$$

$$=-\ln|a^2\sin^2\alpha+d^2|+\ln|y-a\cos\alpha+R(y,a;d,\alpha)|=I11(y;a,d,\alpha)-\ln|a^2\sin^2\alpha+d^2|.$$

Аналитическое представление интеграла (52) имеет вид [14]:

$$I12(y;a,d,\alpha) = \int \frac{dy}{R(y,a;d,\alpha)[a - y\cos\alpha + R(y,a;d,\alpha)]} = \left|\frac{2d}{\sin\alpha}\operatorname{arctg}\left[\frac{a+y+R(y,a;d,\alpha)}{d}\operatorname{tg}\frac{\alpha}{2}\right]\right|. \quad (61)$$

Соотношения (45, 46, 50, 57 и 61) позволяют представить неопределенный интеграл (43) в виде (62). С помощью соотношения (62) не представляет труда вычислить взаимный потенциальный коэффициент двух произвольно расположенных проводников, изображенных на рисунке 1. Поскольку средний потенциал первого проводника, создаваемый равномерным распределением единичного заряда по длине второго, будет $p_{12} = \bar{\varphi}_{12}/Q_2$, то из (42, 43 и 62) после подстановки пределов интегрирования, получим выражение (63).

Нетрудно заметить, что соотношение (63) с учетом (64 и 72), сводится к выражению, приведенному в [14], однако содержит некоторые отличия от приведенного в [3]. Двух независимо полученных результатов свидетельствует о высокой степени их достоверности и возможности практического применения.

Следствием универсальности выражения (63) является его громоздкость, однако при рассмотрении конкретных пар проводников оно, как правило, существенно упрощается. В связи с чем представляется полезным рассмотрение тех частных случаев взаимного расположения, которые встречаются в конструкциях несимметричных вибраторов с вынесенной точкой питания (НВВТП).

$$J(y;a,d,\alpha) = \int \ln|a - y\cos\alpha + R(y,a;d,\alpha)|dy = -y + y\ln|a - y\cos\alpha + R(y,a;d,\alpha)| + a\ln|a\cos\alpha - y + R(y,a;d,\alpha)|dy + \frac{2d}{\sin\alpha}\operatorname{arctg}\left[\frac{a+y+R(y,a;d,\alpha)}{d}\operatorname{tg}\frac{\alpha}{2}\right], \quad (62)$$

$$p_{12} = \frac{1}{4\pi\varepsilon_a l_1 l_2} \times \left\{F1 + F2 + F3 + F4 + \frac{2d}{\sin\alpha}[A1 - A2 - A3 + A4]\right\}, \quad (63)$$

где

$$F1 = (h_1 + l_1)\ln\frac{h_2 + l_2 - (h_1 + l_1)\cos\alpha + R(h_1 + l_1, h_2 + l_2, d, \alpha)}{h_2 - (h_1 + l_1)\cos\alpha + R(h_1 + l_1, h_2, d, \alpha)}, \quad (64)$$

$$F2 = h_1\ln\frac{h_2 - h_1\cos\alpha + R(h_1, h_2, d, \alpha)}{h_2 + l_2 - h_1\cos\alpha + R(h_1, h_2 + l_2, d, \alpha)}, \quad (65)$$

$$F3 = (h_2 + l_2)\ln\frac{h_1 + l_1 - (h_2 + l_2)\cos\alpha + R(h_1 + l_1, h_2 + l_2, d, \alpha)}{h_1 - (h_2 + l_2)\cos\alpha + R(h_1, h_2 + l_2, d, \alpha)}, \quad (66)$$

$$F4 = h_2\ln\frac{h_1 - h_2\cos\alpha + R(h_1, h_2, d, \alpha)}{h_1 + l_1 - h_2\cos\alpha + R(h_1 + l_1, h_2, d, \alpha)}, \quad (67)$$

$$A1 = \operatorname{arctg}\left[\frac{h_2 + l_2 + h_1 + l_1 + R(h_1 + l_1, h_2 + l_2, d, \alpha)}{d}\operatorname{tg}\frac{\alpha}{2}\right], \quad (68)$$

$$A2 = \operatorname{arctg}\left[\frac{h_2 + l_2 + h_1 + R(h_1, h_2 + l_2, d, \alpha)}{d}\operatorname{tg}\frac{\alpha}{2}\right], \quad (69)$$

$$A3 = \operatorname{arctg}\left[\frac{h_2 + h_1 + l_1 + R(h_1 + l_1, h_2, d, \alpha)}{d}\operatorname{tg}\frac{\alpha}{2}\right], \quad (70)$$

$$A4 = \operatorname{arctg}\left[\frac{h_2 + h_1 + R(h_1, h_2, d, \alpha)}{d}\operatorname{tg}\frac{\alpha}{2}\right], \quad (71)$$

$$R(a,b;d,\alpha) = \sqrt{a^2 + b^2 + d^2 - 2ab\cos\alpha} \quad (72)$$

Комплексная емкость несимметричных вибраторов с вынесенной точкой питания

Схематическое изображение НВВТП в общем виде представлено на рисунке 2. По функциональному назначению и конструктивным особенностям все проводники НВВТП можно разделить на группы. При этом для удобства представления обобщенной МВПК (22) в виде клеток, объединяющих проводники по среде размещения, а обобщенной матрицы взаимных емкостей – по принадлежности к верхнему или нижнему плечу НВВТП (23), проводники одной группы не должны находиться в разных средах или принадлежать к разным плечам НВВТП. В результате получим пять групп.

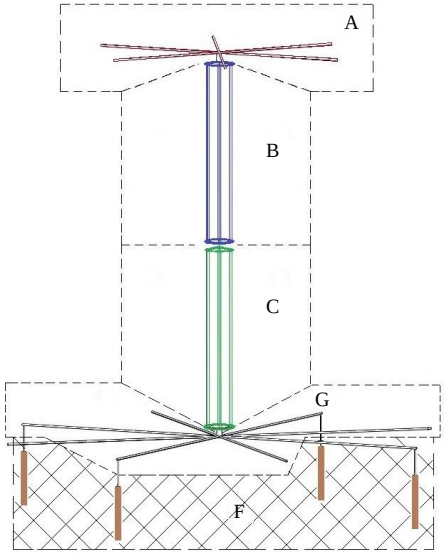


Рис. 2. Схематическое изображение НВВТП в общем виде

Fig. 2. Schematic Representation of the Unbalanced Monopole with Shunt Feed

1) Группа A (верхняя нагрузка) – радиально расходящиеся горизонтальные проводники, соединенные с вершинами вертикальных проводников группы B (предназначены для увеличения действующей длины НВВТП).

2) Группа B (верхнее плечо излучателя) – вертикальные проводники, вершины которой имеют электрический контакт с проводниками группы A (предназначены для связи с внешним электромагнитным полем).

3) Группа C (нижнее плечо излучателя) – вертикальные проводники, основания которых имеют электрический контакт с проводниками группы G (предназначены для связи с внешним электромагнитным полем).

4) Группа G (противовесы) – радиально расходящиеся горизонтальные проводники, имеющие электрический контакт с основаниями проводников группы C и вершинами проводников группы F (предназначены для уменьшения потерь в подстилающей поверхности).

5) Группа F (заземлители) – вертикальные проводники, находящиеся в полупроводящей среде и имеющие электрический контакт с проводниками группы G (предназначены для уменьшения потерь, а также фиксации НВВТП в вертикальном положении).

Можно заметить, что все проводники НВВТП и их изображения при попарном рассмотрении образуют одно из пяти сочетаний, представленных на рисунке 3. На рисунке 3а изображены два непересекающихся проводника, расположенных так, что их средние точки лежат на прямой ортогональной обоим проводникам. Взаимный потенциальный коэффициент таких проводников легко определить из (63), считая, что $h_1 = -l_1/2$, $h_2 = -l_2/2$:

$$p_{12} = \frac{1}{4\pi\varepsilon_a l_1 l_2} \times \times \left\{ \Phi 1 + \Phi 2 + \frac{2d}{\sin\alpha}[A'1 - A'2 - A'3 + A'4] \right\}, \quad (73)$$

где

$$\Phi 1 = l_1 \ln[B(l_2 - l_1, 2d, \alpha) B(l_2, l_1, 2d, \alpha)], \quad (74)$$

$$\Phi 2 = l_2 \ln[B(l_1 - l_2, 2d, \alpha) B(l_1, l_2, 2d, \alpha)], \quad (75)$$

$$B(l_1, l_2, d, \alpha) = = \frac{l_1 + l_2\cos\alpha + \sqrt{{l_1}^2 + {l_2}^2 + d^2 + 2l_1 l_2 \cos\alpha}}{\sqrt{d^2 + {l_1}^2 \sin^2\alpha}}, \quad (76)$$

$$A'1 = \operatorname{arctg}\left[\frac{R(l_1, l_2; 2d, \alpha) + l_1 + l_2}{2d} \operatorname{tg}\frac{\alpha}{2}\right], \quad (77)$$

$$A'2 = \operatorname{arctg}\left[\frac{R(-l_1, l_2; 2d, \alpha) - l_1 + l_2}{2d} \operatorname{tg}\frac{\alpha}{2}\right], \quad (78)$$

$$A'3 = \operatorname{arctg}\left[\frac{R(l_1, -l_2; 2d, \alpha) + l_1 - l_2}{2d} \operatorname{tg}\frac{\alpha}{2}\right], \quad (79)$$

$$A'4 = \operatorname{arctg}\left[\frac{R(-l_1, -l_2; 2d, \alpha) - l_1 - l_2}{2d} \operatorname{tg}\frac{\alpha}{2}\right]. \quad (80)$$

Взаимное расположение одного из проводников верхней нагрузки (любого) и, зеркального изображения другого, представлено на рисунке 3а при условии, что $l_1 = l_2$. Взаимный потенциальный коэффициент этой пары будет вычисляться по выражению (81).

Взаимный потенциальный коэффициент двух параллельных проводников разной длины при условии, что их средние точки лежат на одной прямой, ортогональной обоим проводникам (рисунок 3b), можно получить из (73) при условии, что $\alpha = 0$. Легко заметить, что при этом во втором слагаемом соотношения (73) появляются неопределенности типа 0/0. Раскрывая их по правилу Лопиталя и опуская при этом несложные, но громоздкие преобразования, получим выражение (82). Если же длины проводников равны, то выражение (82) существенно упрощается (83).

Для описания взаимодействия между проводниками верхней нагрузки НВВТП, а также между проводниками, образующими систему противовесов, необходимо найти взаимный потенциальный коэффициент двух скрещенных проводников равной длины, пересекающихся в средней точке (рисунок 3c). Очевидно, что решение этой задачи представляет собой частный случай выражения (81) при $d = 0$. Поскольку разность арктангенсов, стоящая в квадратных скобках второго слагаемого, величина конечная, второе слагаемое тождественно равно нулю. При подстановке $d = 0$ в первое слагаемое, выражение (81) преобразуется к виду (84).

$$p_{12} = \frac{2}{4\pi\varepsilon_a l}\left\{\ln\frac{4\left[l\sin^2\frac{\alpha}{2}+\sqrt{l^2\sin^2\frac{\alpha}{2}+d^2}\right]\left[l\cos^2\frac{\alpha}{2}+\sqrt{l^2\cos^2\frac{\alpha}{2}+d^2}\right]}{l^2\sin^2\alpha+4d^2}+\right.$$

$$\left.+\frac{d}{l\sin\alpha}\left[\operatorname{arctg}\frac{\sqrt{1+\left(\frac{l}{d}\sin\frac{\alpha}{2}\right)^2}}{\operatorname{ctg}\alpha+\left(\frac{l}{2d}\right)^2\sin\alpha}-2\operatorname{arctg}\sqrt{\operatorname{tg}^2\frac{\alpha}{2}+\left(\frac{l\sin\alpha/2}{d}\right)^2}\right]\right\}, \quad (81)$$

$$p_{12} = \frac{1}{4\pi\varepsilon_a l_1 l_2}[l_1\ln B(l_2,-l_1;2d,0)+l_2\ln B(l_1,-l_2;2d,0)+(l_1+l_2)\ln B(l_1,l_2;2d,0)+$$

$$+\sqrt{(l_1-l_2)^2+4d^2}-\sqrt{(l_1+l_2)^2+4d^2}], \quad (82)$$

$$p_{12} = \frac{2}{4\pi\varepsilon_a l}\left[\ln\left(\frac{l}{d}+\sqrt{1+\frac{l^2}{d^2}}\right)+\frac{d}{l}-\sqrt{1+\frac{d^2}{l^2}}\right], \quad (83)$$

$$p_{12} = \frac{2}{4\pi\varepsilon_a l}\ln\left[\left(1+\frac{1}{\sin\alpha/2}\right)\left(1+\frac{1}{\cos\alpha/2}\right)\right]. \quad (84)$$

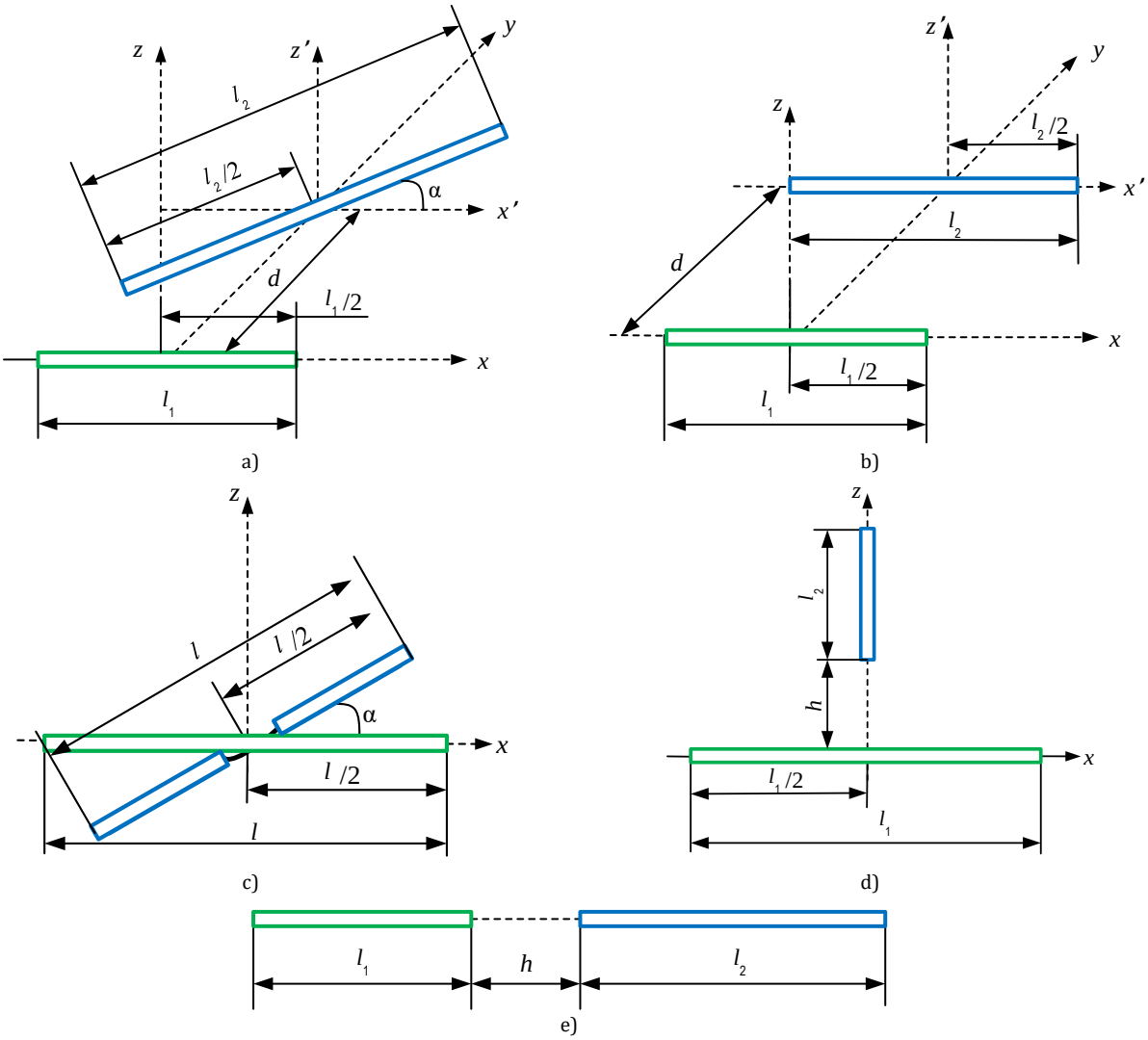


Рис. 3. Возможные пары проводников в конструкции НВВТП

Fig. 3. Possible Pairs of Conductors in the Design of Unbalanced Monopole with Shunt Feed

Формулы (83) и (84) совпадают с приведенными в справочниках [3, 14], что свидетельствует о корректности выполненных преобразований.

Для определения взаимного потенциала вертикальных проводников НВВТП с проводниками верхней нагрузки или системы противовесов (рисунок 3d) следует воспользоваться общим выражением (63), определяющим взаимный потенциал двух произвольных проводников, при следующих условиях: $d = 0$, $h_1 = -l_1/2, h_2 = h$, $\alpha = \pi/2$. Опуская промежуточные преобразования, получим:

$$p_{12} = \frac{1}{4\pi\varepsilon_a l_1 l_2}\left\{ l_1 \ln \frac{2(l_2+h)+\sqrt{{l_1}^2+4(l_2+h)^2}}{2h+\sqrt{{l_1}^2+4h^2}} + \right.$$
$$+2l_2 \ln \frac{l_1+\sqrt{{l_1}^2+4(l_2+h)^2}}{2(l_2+h)} + \qquad (85)$$
$$\left. +2h \ln \frac{h\left[l_1+\sqrt{{l_1}^2+4(l_2+h)^2}\right]}{(l_2+h)\left[l_1+\sqrt{{l_1}^2+4h^2}\right]} \right\}.$$

Для определения взаимного потенциального коэффициента двух коллинеарных проводников (рисунок 3e) можно также воспользоваться соотношением (63), считая при этом $d = 0$, $\alpha = \pi$, $h_1 = 0$, $h_2 = h$. В результате получим (86).

Как известно, в однородной безграничной среде, собственные потенциальные коэффициенты цилиндрических проводников с радиусом a_r и длиной l, при условии, что $l >> a_r$, определяются выражением (87) из [2, 3, 14]. Взаимные же потенциальные коэффициенты двух проводников, ориентация которых может встретиться в конструкции НВВТП, при отсутствии раздела границы сред, определяются формулами (72–81). Эти соотношения совместно с (7, 17–20 и 87) являются достаточными для построения обобщенной матрицы взаимных потенциальных коэффициентов НВВТП и последующего определения его КЕ.

В качестве примера рассмотрим порядок определения КЕ НВВТП, конструкция которого состоит из пяти групп проводников, как показано на рисунке 2. Пусть верхняя нагрузка НВВТП (группа A) содержит A проводников длиной l_A, система противовесов (группа G) – G проводников длиной l_G, плечи излучателя и заземлитель (группы B, C, и F, соответственно) содержат по одному проводнику длиной l_B, l_C и l_F, причем все эти три проводника – коллинеарны. Пусть радиусы проводников групп A, B, C, G и F будут, a_A, a_B, a_C, a_G и a_F, соответственно.

Обобщенную матрицу взаимных потенциальных коэффициентов такого НВВТП удобно представить в клеточной форме, где каждая диагональная клетка представляет собой квадратную матрицу собственных и взаимных потенциальных коэффициентов какой-либо группы проводников. Недиагональные клетки, в общем случае, являются прямоугольными матрицами взаимных потенциальных коэффициентов проводников разных групп. Обобщенная клеточная матрица взаимных потенциальных коэффициентов НВВТП, изображенного на рисунке 2, будет иметь вид (88).

$$p_{12} = \frac{1}{4\pi\varepsilon_a l_1 l_2}\left\{ l_1 \ln \frac{h+l_1+l_2}{h+l_1} + l_2 \ln \frac{h+l_1+l_2}{h+l_2} + h \ln \frac{h(h+l_1+l_2)}{(h+l_1)(h+l_2)} \right\}, \qquad (86)$$

$$p_{11} = \frac{2}{4\pi\varepsilon_a l}\left(\ln \frac{2l}{a_r} - 1\right), \qquad (87)$$

$$\left|P^{\Sigma}_{K\,K}\right| = \begin{vmatrix} \left|P^{\Sigma}_{AA}\right| & \left|P^{\Sigma}_{AB}\right| & \left|P^{\Sigma}_{AC}\right| & \left|P^{\Sigma}_{AG}\right| & \left|P^{\Sigma}_{A\underline{F}}\right| \\ \left|P^{\Sigma}_{BA}\right| & \left|P^{\Sigma}_{BB}\right| & \left|P^{\Sigma}_{BC}\right| & \left|P^{\Sigma}_{BG}\right| & \left|P^{\Sigma}_{B\underline{F}}\right| \\ \left|P^{\Sigma}_{CA}\right| & \left|P^{\Sigma}_{CB}\right| & \left|P^{\Sigma}_{CC}\right| & \left|P^{\Sigma}_{CG}\right| & \left|P^{\Sigma}_{C\underline{F}}\right| \\ \left|P^{\Sigma}_{GA}\right| & \left|P^{\Sigma}_{GB}\right| & \left|P^{\Sigma}_{GC}\right| & \left|P^{\Sigma}_{G\bar{G}}\right| & \left|P^{\Sigma}_{G\underline{F}}\right| \\ \left|P^{\Sigma}_{\underline{F}A}\right| & \left|P^{\Sigma}_{\underline{F}B}\right| & \left|P^{\Sigma}_{\underline{F}C}\right| & \left|P^{\Sigma}_{\underline{F}G}\right| & \left|P^{\Sigma}_{\underline{F}\underline{F}}\right| \end{vmatrix} = \begin{vmatrix} \left|P^{\Sigma}_{\bar{N}\,\bar{N}}\right| & \left|P^{\Sigma}_{\bar{N}\,\underline{M}}\right| \\ \left|P^{\Sigma}_{\underline{M}\,\bar{N}}\right| & \left|P^{\Sigma}_{\underline{M}\,\underline{M}}\right| \end{vmatrix}. \qquad (88)$$

Элементы квадратной матрицы $\left|P^{\Sigma}_{AA}\right|$ являются собственными и взаимными потенциальными коэффициентами проводников верхней нагрузки, и, как следует из (17), имеют вид:

$$p^{\Sigma}_{\overline{a}\overline{a'}} = p_{\overline{a}\overline{a'}} + \gamma p_{\overline{a}\,\underline{a'}} \qquad (89)$$

где a, a' – номера проводников группы A $(1 \le a \le A;\ 1 \le a' \le A)$; $p_{\bar{a}\,\bar{a'}}$ – взаимный потенциальный коэффициент проводников a и a', размещенных в безграничной однородной среде с параметрами воздуха; $p_{\bar{a}\,\underline{a'}}$ – взаимный потенциальный коэффициент проводника a и зеркального изображения проводника a', при условии размещения их в безграничной однородной среде с параметрами воздуха.

Если $a \ne a'$, то $p_{\bar{a}\,\bar{a'}}$ определяется соотношением (84) при $l = l_A; \varepsilon_a = \varepsilon_0;\ \alpha = \alpha_{a\,a'} = (\pi/A)(a - a')$;

$p_{\bar{a}\ \underline{a'}}$ определяется соотношением (81) при: $l = l_A$; $\varepsilon_a = \varepsilon_0$; $\alpha = \alpha_{a\,a'} = (\pi/\mathrm{A})(a - a')$; $d = l_B + l_C$.

При $a = a'$: $p_{\bar{a}\ \bar{a}}$ – собственный потенциальный коэффициент проводника a в безграничной среде, определяемый соотношением (87) при $l = l_A$; $\varepsilon_a = \varepsilon_0$; $a_r = a_A$; $p_{\bar{a}\ \underline{a}}$ – взаимный потенциальный коэффициент проводника a и его изображения при условии их нахождения в безграничной среде, определяемый соотношением (83) при $l = l_A$; $\varepsilon_a = \varepsilon_0$; $d = l_B + l_C$.

Матрица $\left|P^{\Sigma}_{\overline{BB}}\right|$ состоит из одного элемента, являющегося собственным потенциальным коэффициентом верхнего плеча излучателя НВВТП. Из соотношения (17) следует:

$$p^{\Sigma}_{\overline{bb}} = p_{\overline{bb}} + \gamma p_{\overline{b}\ \underline{b}}, \qquad (90)$$

где $p_{\overline{b}\ \overline{b}}$ – определяется соотношением (87) при $l = l_B$; $\varepsilon_a = \varepsilon_0$; $a_r = a_B$; $p_{\overline{b}\ \underline{b}}$ – соотношением (86) при $l_1 = l_2 = l_B$; $\varepsilon_a = \varepsilon_0$; $h = 2l_C$.

Матрица $\left|P^{\Sigma}_{\overline{CC}}\right|$ также состоит из одного элемента, являющегося собственным потенциальным коэффициентом нижнего плеча излучателя НВВТП. Из соотношения (17) следует:

$$p^{\Sigma}_{\overline{c}\overline{c}} = p_{\overline{c}\ \overline{c}} + \gamma p_{\overline{c}\ \underline{c}}, \qquad (91)$$

где $p_{\overline{c}\ \overline{c}}$ – определяется соотношением (87) при $l = l_C$; $\varepsilon_a = \varepsilon_0$; $a_r = a_C$; $p_{\overline{c}\ \underline{c}}$ – соотношением (86) при $l_1 = l_2 = l_C$; $\varepsilon_a = \varepsilon_0$; $h = 0$.

Элементы квадратной матрицы $\left|P^{\Sigma}_{\overline{G}\overline{G}}\right|$ являются собственными и взаимными потенциальными коэффициентами проводников системы противовесов. В общем виде они также определяются соотношением (17):

$$p^{\Sigma}_{\bar{g}\ g'} = p_{\overline{g}\ \overline{g'}} + jp_{\bar{g}\underline{g'}}, \qquad (92)$$

при условии, что $1 \le g \le G; 1 \le g' \le G$.

Если $g \ne g'$, то $p_{\overline{g}\ \overline{g'}}$ – определяется соотношением (84) при $l = l_G; \varepsilon_a = \varepsilon_0$; $\alpha = \alpha_{g\,g'} = (\pi/\mathrm{G})(g - g')$; $p_{\bar{g}\underline{g'}}$ – соотношением (81) при $l = l_G$; $\varepsilon_a = \varepsilon_0$; $\alpha = \alpha_{g\,g'} = (\pi/\mathrm{G})(g - g')$; $d = 2a_D$.

Если $g = g'$, то: $p_{\overline{g}\ \overline{g'}}$ – определяется соотношением (87) при $l = l_G; \varepsilon_a = \varepsilon_0; a_r = a_G; p_{\bar{g}\underline{g'}}$ – соотношением (83) при $l = l_G; \varepsilon_a = \varepsilon_0; d = 2a_G$.

Прямоугольная матрица $\left|P^{\Sigma}_{\overline{AB}}\right|$ представляет собой столбец, содержащий A элементов, являющихся взаимными потенциальными коэффициентами проводников a и b. В силу осесимметричности расположения проводников относительно проводника b все элементы столбца равны:

$$p^{\Sigma}_{\bar{a}\,\bar{b}} = p_{\bar{a}\,\bar{b}} + \gamma p_{\bar{a}\,\underline{b}}, \qquad (93)$$

где $p_{\bar{a}\,\bar{b}}$ – определяется соотношением (85) при $l_1 = 2l_A; \varepsilon_a = \varepsilon_0; l_2 = l_B; h = 0$; $p_{\bar{a}\,\underline{b}}$ – соотношением (85) при $l_1 = 2l_A; \varepsilon_a = \varepsilon_0; l_2 = l_B; h = 2l_C$.

Прямоугольная матрица $\left|P^{\Sigma}_{\overline{AC}}\right|$ также представляет собой столбец, содержащий A одинаковых элементов, представляющих собой взаимные потенциальные коэффициенты проводников a с проводником c, как следует из соотношения (17):

$$p^{\Sigma}_{\bar{a}\,\bar{c}} = p_{\bar{a}\,\bar{c}} + \gamma p_{\bar{a}\underline{c}}, \qquad (94)$$

где $p_{\bar{a}\,\bar{c}}$ – определяется соотношением (85) при $l_1 = 2l_A; \varepsilon_a = \varepsilon_0; l_2 = l_C; h = l_B$; $p_{\overline{\bar{a}}\ \underline{c}}$ – соотношением (85) при $l_1 = 2l_A; \varepsilon_a = \varepsilon_0; l_2 = l_C; h = l_B + l_C$.

Прямоугольная матрица $\left|P^{\Sigma}_{\overline{AG}}\right|$ содержит A строк и G столбцов, состоящих из взаимных потенциальных коэффициентов проводников группы A с проводниками группы G. Общий вид элементов этой матрицы также определяется соотношением (17):

$$p^{\Sigma}_{\bar{a}\,\bar{g}} = p_{\bar{a}\,\bar{g}} + \gamma p_{\bar{a}\,\underline{g}}. \qquad (95)$$

Легко показать, что угол между проекцией проводника g на плоскость проводников группы A составляет:

$$\alpha(a\ g) = \pi/2\ \ [(a/\mathrm{A}) - ([g/G])]. \qquad (96)$$

Если $\alpha_{a\,g} \ne 0$, то $p_{\bar{a}\,\bar{g}}$ – определяется соотношением (73) при $l_1 = 2l_A; \varepsilon_a = \varepsilon_0; l_2 = 2l_G; d = l_B + l_C$.

Если $\alpha_{a\,g} = 0$, то: $p_{\bar{a}\,\bar{g}}$ – соотношением (82) при тех же условиях. В обоих случаях $p_{\bar{a}\,\underline{g}} = p_{\bar{a}\,\bar{g}}$.

Для рассматриваемой конструкции НВВТП, прямоугольная матрица $\left|P^{\Sigma}_{\overline{BC}}\right|$ состоит из одного элемента, определяемого в соответствии с (17) как:

$$p^{\Sigma}_{\bar{b}\,\bar{c}} = p_{\bar{b}\,\bar{c}} + \gamma p_{\bar{b}\,\underline{c}}, \qquad (97)$$

где $p_{\bar{b}\,\bar{c}}$ – определяется соотношением (86) при $l_1 = l_B; \varepsilon_a = \varepsilon_0; l_2 = l_C; h = 0$; $p_{\bar{b}\,\underline{c}}$ – соотношением (86) при $l_1 = l_B; \varepsilon_a = \varepsilon_0; l_2 = l_C; h = l_C$.

Прямоугольная матрица $\left|P^{\Sigma}_{\overline{BG}}\right|$ является матрицей-строкой, состоящей из G одинаковых элементов:

$$p^{\Sigma}_{\bar{b}\,\bar{g}} = p_{\bar{b}\,\bar{g}} + \gamma p_{\bar{b}\,\underline{g}}, \qquad (98)$$

где $p_{\bar{b}\,\bar{g}} = p_{\bar{b}\,\underline{g}}$ – определяется соотношением (85) при $l_1 = 2l_G; \varepsilon_a = \varepsilon_0; l_2 = l_B; h = l_C$.

Прямоугольная матрица $\left|P^{\Sigma}_{\overline{CG}}\right|$ также является строкой, состоящей из G одинаковых элементов:

$$p^{\Sigma}_{\bar{c}\,\bar{g}} = p_{\bar{c}\,\bar{g}} + \gamma p_{\bar{c}\,\underline{g}}, \qquad (99)$$

где $p_{\bar{c}\,\bar{g}} = p_{\bar{c}\,\underline{g}}$ и определяется соотношением (85) при $l_1 = 2l_C; \varepsilon_a = \varepsilon_0; l_2 = l_C; h = 0$.

Учитывая, что согласно принципу взаимности $\left|P^{\Sigma}_{\overline{BA}}\right| = \left|P^{\Sigma}_{\overline{AB}}\right|^{T}$, $\left|P^{\Sigma}_{\overline{CA}}\right| = \left|P^{\Sigma}_{\overline{AC}}\right|^{T}$, $\left|P^{\Sigma}_{\overline{CB}}\right| = \left|P^{\Sigma}_{\overline{BC}}\right|^{T}$, $\left|P^{\Sigma}_{\overline{GA}}\right| = \left|P^{\Sigma}_{\overline{AG}}\right|^{T}$, $\left|P^{\Sigma}_{\overline{GB}}\right| = \left|P^{\Sigma}_{\overline{BG}}\right|^{T}$ и $\left|P^{\Sigma}_{\overline{GC}}\right| = \left|P^{\Sigma}_{\overline{CG}}\right|^{T}$, все элементы обобщенной квадратной матрицы $\left|P^{\Sigma}_{\overline{NN}}\right|$, содержащей собственные и взаимные потенциальные коэффициенты проводников, находящихся в воздушной среде, могут быть вычислены по изложенному выше алгоритму.

Поскольку группа F состоит из одного проводника, прямоугольные матрицы $\left|P^{\Sigma}_{\overline{A}\underline{F}}\right|$ и $\left|P^{\Sigma}_{\overline{G}\underline{F}}\right|$ представляют собой столбцы, содержащие A и G элементов, соответственно. При этом каждая из указанных матриц состоит из одинаковых элементов, так как проводники групп A и G имеют одинаковое расположение относительно проводника F. Эти элементы представляют собой потенциалы, создаваемые на проводниках групп A и G единичным зарядом, распределенным по проводнику F, при условии, что все проводники находятся в однородной среде с параметрами воздуха. Согласно соотношению (18), эти потенциалы будут следующими:

$$p^{\Sigma}_{\overline{a}\,\underline{f}} = (1+\gamma)p_{\overline{a}\,\underline{f}}, \qquad (100)$$

$$p^{\Sigma}_{\overline{g}\,\underline{f}} = (1+\gamma)p_{\overline{g}\,\underline{f}}, \qquad (101)$$

где $p_{\overline{a}\,\underline{f}}$ – определяется соотношением (85) при $l_1 = 2l_A; \varepsilon_a = \varepsilon_0; l_2 = l_F; h = l_B + l_C$; $p_{\overline{g}\,\underline{f}}$ – определяется соотношением (85) при $l_1 = 2l_G; \varepsilon_a = \varepsilon_0$; $l_2 = l_F; h = 0$.

Матрицы $\left|P^{\Sigma}_{\overline{B}\underline{F}}\right|$ и $\left|P^{\Sigma}_{\overline{C}\underline{F}}\right|$ содержат по одному потенциальному коэффициенту; их общий вид определяется аналогичным образом:

$$p^{\Sigma}_{\overline{b}\,\underline{f}} = (1+\gamma)p_{\overline{b}\,\underline{f}}, \qquad (102)$$

$$p^{\Sigma}_{\overline{c}\,\underline{f}} = (1+\gamma)\,p_{\overline{c}\,\underline{f}}. \qquad (103)$$

Поскольку проводники $\overline{b}$, $\overline{c}$ и $\underline{f}$ коллинеарны, то $p_{\overline{b}\,\underline{f}}$ – определяется соотношением (86) при $l_1 = l_B; \varepsilon_a = \varepsilon_0; l_2 = l_F; h = l_C$; $p_{\overline{c}\,\underline{f}}$ – определяется соотношением (86) при $l_1 = l_C; \varepsilon_a = \varepsilon_0; l_2 = l_F; h = 0$.

Легко показать, что матрицы строки $\left|P^{\Sigma}_{\underline{F}\overline{A}}\right|$, $\left|P^{\Sigma}_{\underline{F}\overline{B}}\right|$, $\left|P^{\Sigma}_{\underline{F}\overline{C}}\right|$ и $\left|P^{\Sigma}_{\underline{F}\overline{G}}\right|$, определенные в соответствии с (19), удовлетворяют соотношениям:

$$\left|P^{\Sigma}_{\underline{F}\overline{A}}\right| = \left|P^{\Sigma}_{\overline{A}\underline{F}}\right|^{T}; \left|P^{\Sigma}_{\underline{F}\overline{G}}\right| = \left|P^{\Sigma}_{\overline{G}\underline{F}}\right|^{T};$$
$$\left|P^{\Sigma}_{\underline{F}\overline{B}}\right| = \left|P^{\Sigma}_{\overline{B}\underline{F}}\right|^{T}; \left|P^{\Sigma}_{\underline{F}\overline{C}}\right| = \left|P^{\Sigma}_{\overline{C}\underline{F}}\right|^{T}. \qquad (104)$$

Последней из рассмотренной матрицей клеткой, входящей в обобщенную матрицу взаимных потенциальных коэффициентов НВВТП (88), является $\left|P^{\Sigma}_{\underline{F}\underline{F}}\right|$, описывающая взаимодействие проводников, заглубленных в землю, а именно проводников группы F. Поскольку в анализируемой конструкции эта группа состоит из одного проводника, матрица $\left|P^{\Sigma}_{\underline{F}\underline{F}}\right|$ состоит из одного элемента, общий вид которого, в соответствии с (20), будет следующим:

$$p^{\Sigma}_{\underline{f}\,\underline{f}} = p_{\underline{f}\,\underline{f}} - \gamma p_{\underline{f}\,\bar{f}}, \qquad (105)$$

где $p_{\underline{f}\,\underline{f}}$ – собственный потенциальный коэффициент проводника f в безграничной среде с параметрами земли, определяемый соотношением (87) при $l = l_F; \varepsilon_a = \varepsilon_0\,\varepsilon_k; a_r = a_F$; $p_{\underline{f}\,\bar{f}}$ – взаимный потенциальный коэффициент проводника и его изображения при условии, что они находятся в среде с параметрами земли. Определяется соотношением (86) при $l_1 = l_F; l_2 = l_F; \varepsilon_a = \varepsilon_0\,\varepsilon_k$.

Соотношения (89–105), совместно с (73–87) полностью определяют обобщенную матрицу взаимных потенциальных коэффициентов проводников НВВТП, анализируемой конструкции. В результате обращения эта матрица преобразуется в матрицу взаимных емкостей, в которой можно выделить клетки, описывающие взаимодействие проводников по принципу принадлежности к верхнему или нижнему плечу НВВТП:

$$\left|P^{\Sigma}_{K\,K}\right|^{-1} = \begin{vmatrix} |C_{\overline{AA}}| & |C_{\overline{AB}}| & |C_{\overline{AC}}| & |C_{\overline{AG}}| & \left|C_{\overline{A}\underline{F}}\right| \\ |C_{\overline{BA}}| & |C_{\overline{BB}}| & |C_{\overline{BC}}| & |C_{\overline{BG}}| & \left|C_{\overline{B}\underline{F}}\right| \\ |C_{\overline{CA}}| & |C_{\overline{CB}}| & |C_{\overline{CC}}| & |C_{\overline{CG}}| & \left|C_{\overline{C}\underline{F}}\right| \\ |C_{\overline{GA}}| & |C_{\overline{GB}}| & |C_{\overline{GC}}| & |C_{\overline{G}\overline{G}}| & \left|C_{\overline{G}\underline{F}}\right| \\ \left|C_{\underline{F}\overline{A}}\right| & \left|C_{\underline{F}\overline{B}}\right| & \left|C_{\underline{F}\overline{C}}\right| & |C_{\underline{F}\overline{G}}| & |C_{\underline{F}\underline{F}}| \end{vmatrix} = \begin{vmatrix} |C_{VV}| & |C_{VT}| \\ |C_{TV}| & |C_{TT}| \end{vmatrix}, \qquad (106)$$

где $V = A + B$ – число проводников, содержащихся в верхнем плече НВВТП; $T = C + G + F$ – число проводников, содержащихся в нижнем плече НВВТП.

Определив матрицы $|C_{VV}|$, $|C_{VT}|$, $|C_{TV}|$ и $|C_{TT}|$, суммированием их элементов (30–33), легко вычислить комплексные емкости каждого плеча антенны, и взаимные емкости между плечами, после чего, воспользовавшись соотношением (34), найти комплексную емкость НВВТП.

Как известно [2, 3, 10], мнимая составляющая комплексной емкости антенны характеризует сопротивление тепловых потерь в подстилающей поверхности, что позволяет использовать ее в качестве критерия, как при оценке эффективности различных элементов конструкции НВВТП, так и при выборе их размеров.

В работе [10] сопротивление тепловых потерь использовалось как критерий при выборе длины

заземлителя НВВТП ШТ4Н81, состоящего из трех коллинеарных проводников. Сделан вывод о невозможности существенного снижения потерь, путем удлинения заземлителя, при развертывании ШТ4Н81 на сухой почве. Рекомендовано исследовать вопрос о целесообразности введения в конструкцию ШТ4Н81 системы радиально расходящихся противовесов.

На рисунке 4 приведены частотные зависимости сопротивления потерь ШТ4Н81, конструкция которого дополнена верхней нагрузкой, содержащей шесть радиально расходящихся проводников, и шестью, также радиально расходящимися, противовесами. Длина излучателей $l_1 = l_2 = 1$ м. Диаметр проводников: излучателя – 50 мм; верхней нагрузки – 4 мм; противовесов – 2 мм. Длина заземлителя: 0,4 м.

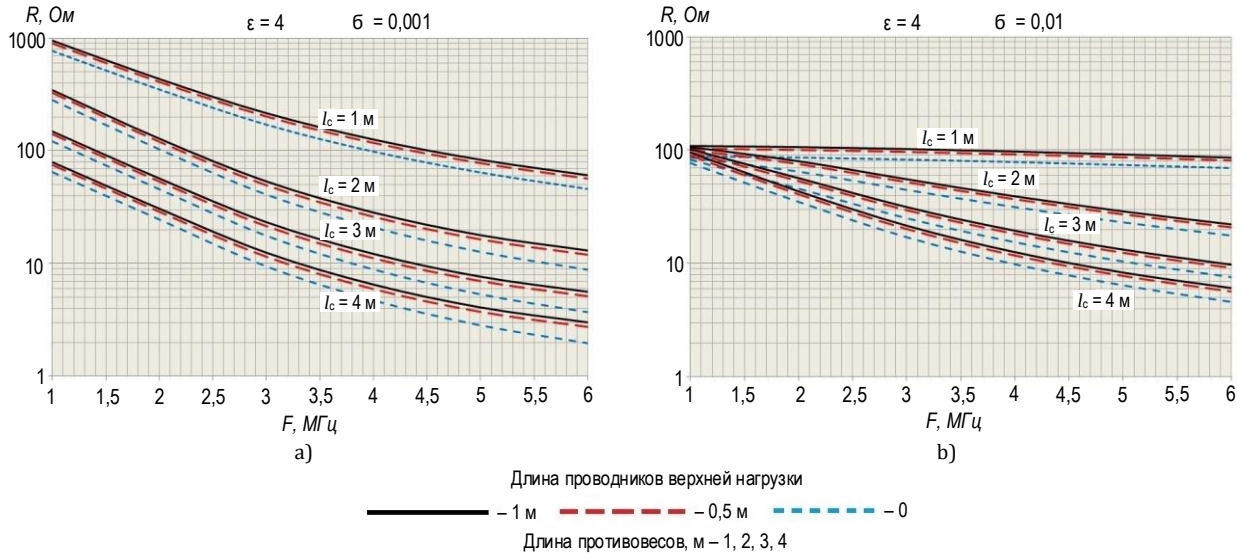


Рис. 4. Вещественная составляющая сопротивления потерь НВВТП в зависимости от частоты для влажной (a) и сухой (b) почв

Fig. 4. The Frequency Dependence of Real Part of Impedance of Unbalanced Monopole with Shunt Feed for Wet (a) and Dry (b) Soils

Как видно из рисунков, введение верхней нагрузки увеличивает сопротивление потерь, что объясняется большей разветвленностью токов смещения. Таким образом, вопрос о целесообразности применения верхней нагрузки можно решить только после оценки ее влияния на сопротивление излучения, а, следовательно, и на КПД НВВТП. Введение в конструкцию системы противовесов НВВТП уменьшает потери, что особенно заметно при развертывании на сухой почве.

Снижение потерь на влажной почве не так значительно, а поскольку развертывание противовесов существенно снижает мобильность комплекса, целесообразно рассмотреть вопрос о проектировании быстросъемной системы противовесов. Это позволит принимать решение об ее использовании руководителю подразделения, эксплуатирующего комплекс, на основании анализа состояния почвы и складывающейся оперативной обстановки.

Заключение

Полученные аналитические выражения позволяют определять КЕ нормально разомкнутых проволочных антенн практически любой конструкции, могут использоваться при решении задач оптимизации размеров их элементов. Предложенная последовательность вычислений при определении КЕ многоэлементных проволочных антенн может быть полезна при определении таких параметров, как входное сопротивление, коэффициенты полезного действия, согласования по сопротивлению и усиления антенны. Приведенный анализ влияния верхней нагрузки и системы противовесов на тепловые потери НВВТП позволил выработать рекомендации по конструированию НВВТП малых по сравнению с длиной волны.

Список источников

1. Попов О.В., Тумашов А.В., Борисов Г.Н., Коровин К.О. Математическая модель несимметричного вибратора с вынесенной точкой питания. Часть 1. Общий подход к построению математической модели // Труды учебных заведений связи. 2023. Т. 9. № 1. С. 24–33. DOI:10.31854/1813-324X-2023-9-1-24-33.
2. Гавеля Н.П., Истрашкин А.Д., Муравьев Ю.К., Серков В.П. Антенны. Ч. I. Л.: ВКАС, 1963. 633 с.
3. Муравьев Ю.К. Справочник по расчету проволочных антенн. Л.: ВАС, 1978.
4. Серков В.П. Распространение радиоволн и антенные устройства. Л.: ВАС, 1981.
5. Зернов Н.В., Карпов В.Г. Теория радиотехнических цепей. Л.: Энергия: Ленингр. отд-ние, 1972. 816 с.
6. Конторович М.И. О расчете емкости антенны по методу Хоу // Труды ВКАС. 1943. № 2.

7. Howe C.W. On the capacity of radio-telegraphic antennae // Electrician. 1914. Vol. 73.
8. Бесчастнов Н.С., Конторович М.И. О потерях в земле при использовании корпуса передатчика в качестве противовеса // Труды ВКАС. 1944. № 3.
9. Конторович М.И. Эквивалентные параметры провода // Труды ВКАС. 1944. № 6.
10. Попов О.В., Тумашов А.В., Борисов Г.Н. Методика расчета сопротивления потерь заземленных несимметричных вибраторов с вынесенной точкой питания // Успехи современной радиоэлектроники. 2021. Т. 75. № 4. С. 71–79. DOI:10.18127/j20700784-202104-10
11. Русин Ю.С. Метод приближенного расчета электрической емкости // Электричество. 1960. № 11. С. 48.
12. Гольдштейн Л.Д., Зернов Н.В. Электромагнитные поля и волны. М.: Изд-во «Советское радио», 1971.
13. Прудников А.П., Брычков Ю.А., Маричев О.И. Интегралы и ряды. Элементарные функции. М.: Наука. Главная редакция физико-математической литературы. 1981.
14. Иоссель Ю.Я., Кочанов Э.С., Струнский М.Г. Расчет электрической емкости. Л.: Энергоиздат: Ленингр. отд-ние, 1981. 288 с.

References

1. Popov O., Tumashov A., Borisov G., Korovin K. Mathematical Model of the Unbalanced Monopole Feed. Part 1. General Approach to Building a Mathematical Model. *Proc. of Telecom. Universities*. 2023;9(1):24–33. (in Russ.) DOI:10.31854/1813-324X-2023-9-1-24-33
2. Gavelya N.P., Istrashkin A.D., Muravyov Yu.K. *Antennas. Part I.* Leningrad: Military Academy of Communications Publ.; 1963. 633 p. (in Russ.)
3. Muravyov Yu.K. *Handbook for the Calculation of Wire Antennas.* Leningrad: Military Academy of Telecommunications Publ.; 1978. (in Russ.)
4. Serkov V.P. *Radio Wave Propagation and Antenna Devices*. Leningrad: Military Academy of Telecommunications Publ.; 1981. (in Russ.)
5. Zernov N.V., Karpov V.G. *Theory of Radio Circuits*. Leningrad: Energiia Publ.; 1972. 816 p. (in Russ.)
6. Kontorovich M.I. On the calculation of the capacitance of the antenna by the Howe method. *Trudy VKAS*. 1944;2. (in Russ.)
7. Howe C.W. On the capacity of radio-telegraphic antennae. *Electrician*. 1914:73.
8. Beschastnov N.S., Kontorovich M.I. On Loss in the Ground when Assembling the Mechanism as a Counterweight. *Trudy VKAS*. 1944;3. (in Russ.)
9. Kontorovich M.I. Equivalent wire parameters. *Trudy VKAS.* 1944;6. (in Russ.)
10. Popov O.V., Tumashov A.V., Borisov G.N. Method for Calculating the Loss Asymmetric Vibrators with a Remote Power Point. *Journal Achievements of Modern Radioelectronics*. 2021;75(4):71–79. DOI:10.18127/j20700784-202104-10
11. Rusin Yu.S. Approximate Capacitance Measurement Method. *Elektrichestvo*. 1960;11(48) (in Russ.)
12. Goldstein L.D., Zernov N.V. *Electromagnetic Fields and Waves*. Moscow: Sovetskoe radio Publ.; 1971. (in Russ.)
13. Prudnikov A.P., Brychkov Yu.A., Marichev O.I. *Integrals and Series. Elementary Functions*. Moscow: Nauka Glavnaia redaktsiia fiziko-matematicheskoi literatury Publ.; 1981. (in Russ.)
14. Iossel Yu.Ya., Kochanov E.S., Strunsky M.G. *Capacity Calculation*. Leningrad: Energoizdat Publ.; 1981. 288 p. (in Russ.)

Статья поступила в редакцию 17.04.2023; одобрена после рецензирования 27.04.2023; принята к публикации 11.05.2023.
The article was submitted 17.04.2023; approved after reviewing 27.04.2023; accepted for publication 11.05.2023.

Информация об авторах:

ПОПОВ Олег Вениаминович — кандидат технических наук, доцент, ведущий научный сотрудник ООО «Специальный технологический центр»
https://orcid.org/0000-0002-5315-2679

ТУМАШОВ Андрей Витальевич — инженер-конструктор ООО «Специальный технологический центр»
https://orcid.org/0000-0003-2656-0463

БОРИСОВ Георгий Николаевич — инженер ООО «Специальный технологический центр»
https://orcid.org/0000-0002-3275-251X

КОРОВИН Константин Олегович — кандидат физико-математических наук, доцент, заведующий кафедрой радиосистем и обработки сигналов Санкт-Петербургского государственного университета телекоммуникаций им. проф. М.А. Бонч-Бруевича
https://orcid.org/0000-0001-7979-3725

DOI:10.31854/1813-324X-2023-9-2

Научная статья
УДК 621.396
DOI:10.31854/1813-324X-2023-9-2-23-39

Прототип приемо-передающего оборудования скоростной передачи данных в частотном диапазоне 57–64 ГГц

Олеся Викторовна Болховская, obol@rf.un.ru
Григорий Александрович Ермолаев, gregory.a.ermolaev@gmail.com
Сергей Николаевич Трушков, trushkovsn@gmail.com
Александр Александрович Мальцев, maltsev@rf.un.ru

Национальный исследовательский Нижегородский государственный университет им. Н.И. Лобачевского, Нижний Новгород, 603950, Российская Федерация

Аннотация: *Целью настоящей работы является создание и исследование характеристик прототипа приемо-передающего оборудования с программно-определяемым функционалом, работающего в миллиметровом диапазоне длин волн в сетях скоростной передачи данных. В ходе работы были решены задачи разработки и программной реализации алгоритмов цифровой обработки сигналов и аппаратной части, проведены экспериментальные измерения характеристик и полевые испытания прототипа. Экспериментальные исследования показали, что разработанное оборудование осуществляет передачу и прием сигналов в диапазоне частот 57–64 ГГц с возможностью дискретного изменения полосы частот сигналов: 100, 200, 400, 800 МГц и поддерживает 12 сигнально-кодовых конструкций с применением кодов с малой плотностью проверки на четность. Применение адаптивного алгоритма демодуляции и декодирования в радиоприемнике позволило повысить эффективность передачи сигналов и уменьшить вероятность пакетных ошибок в два раза. Разработанный прототип обеспечивает скорость передачи данных в пакете 2 Гбит/с на расстояниях до 100 м и 500 Мбит/с на расстояниях до 300 м.*

Ключевые слова: *системы радиосвязи, миллиметровый диапазон длин волн, скоростная передача данных, приемо-передающее оборудование, алгоритмы цифровой обработки сигналов*

Финансирование: Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 20-32-90197.

Ссылка для цитирования: Болховская О.В., Ермолаев Г.А., Трушков С.Н., Мальцев А.А. Прототип приемо-передающего оборудования скоростной передачи данных в частотном диапазоне 57–64 ГГц // Труды учебных заведений связи. 2023. Т. 9. № 2. С. 23–39. DOI:10.31854/1813-324X-2023-9-2-23-39

Prototype of High-Speed Data Transmission Receiving and Transmitting Equipment in the 57–64 GHz Frequency Range

Olesya Bolkhovskaya, obol@rf.un.ru
Gregory Ermolaev, gregory.a.ermolaev@gmail.com
Sergey Trushkov, trushkovsn@gmail.com
Alexander Maltsev, maltsev@rf.un.ru

Lobachevsky State University of Nizhny Novgorod,
Nizhny Novgorod, 603950, Russian Federation

Abstract: *The purpose of this work is to create and study the characteristics of a prototype of receiving and transmitting equipment operating in the millimeter wavelength range in high-speed data transmission networks. During the work, the task of developing and software implementation of digital signal processing algorithms was solved, the hardware part was developed and implemented, experimental measurements of characteristics and field tests of the prototype were carried out. Experimental studies have shown that the developed equipment transmits and receives signals in the frequency range 57–64 GHz with the possibility of discrete change of the signal frequency bandwidth: 100, 200, 400, 800 MHz and supports 12 modulation and coding schemes with low-density parity check code. The use of an adaptive algorithm for demodulation and decoding at the receiver made it possible to increase the efficiency of signal transmission and reduce the probability of packet errors by half. The developed prototype provides a data transmission rate of 2 Gbit/s at distances up to 100 m and of 500 Mbit/s at distances up to 300 m.*

Keywords: *radio communication systems, millimeter wavelength range, high-speed data transmission, receiving and transmitting equipment, digital signal processing algorithms*

Funding: This research was funded by RFBR according to the research project No. 20-32-90197.

For citation: Bolkhovskaya O., Ermolaev G., Trushkov S., Maltsev A. Prototype of High-Speed Data Transmission Receiving and Transmitting Equipment in the 57–64 GHz Frequency Range. *Proc. of Telecom. Universities.* 2023;9(2):23–39. (in Russ.) DOI:10.31854/1813-324X-2023-9-2-23-39

Введение

Мобильные системы связи 5-го поколения (5G) должны обеспечивать передачу данных со скоростями в несколько гигабит в секунду. Такие скорости могут быть достигнуты путем перехода систем связи в миллиметровый диапазон длин волн с использованием существенно более широкополосных сигналов и особой гетерогенной архитектуры сети [1–3]. До недавнего времени миллиметровый диапазон практически не использовался в мобильных системах связи из-за отсутствия доступной элементной базы, необходимой для создания относительно дешевых средств генерации, приема и обработки сигналов. Однако развитие полупроводниковых технологий и прогресс в области изготовления радиочастотных (РЧ) интегральных схем обеспечили возможность серийного производства радиокомпонент миллиметрового диапазона с рабочей частотой 60 ГГц и выше [4]. Появление дешевых и компактных приемопередатчиков сделало миллиметровый диапазон привлекательным для создания новых беспроводных систем связи.

Следует также отметить, что в миллиметровом диапазоне есть малоиспользуемые современными радиоэлектронными средствами участки спектра. Так, в диапазоне частот 57–64 ГГц наблюдается сильное затухание радиоволн (до 15 дБ на 1 км), обусловленное резонансным поглощением излучения молекулами кислорода, спектральные линии которых находятся в окрестности частоты 60 ГГц. Это делает данный диапазон малопригодным для беспроводной передачи данных на большие расстояния (более километра). Поэтому во многих странах мира диапазон 57–64 ГГц является нелицензионным или существуют упрощенные процедуры оформления разрешительных документов на его практическое использование. Однако резонансное поглощение радиоволн в этом диапазоне имеет и положительной эффект – оно слабо влияет на работу систем радиосвязи с радиусом действия порядка 100–200 м и, в то же время, существенно уменьшает взаимную интерференцию между станциями, расположенными на больших расстояниях друг от друга. Это сделало данный диапазон весьма привлекательным для создания новых локальных высокоскоростных систем беспроводного доступа в Интернет.

Уже в 2012 г. комитетом по стандартизации IEEE (Institute of Electrical and Electronics Engineers) был принят первый стандарт миллиметрового диапазона длин волн IEEE 802.11ad [5–7], предназначенный для систем Wi-Fi, работающих в частотном диапазоне 57–64 ГГц. В дальнейшем частотный диапазон и возможности этого стандарта были существенно расширены в принятом в июне 2021 г. стандарте IEEE802.11ay [8–10].

Параллельно проводилось освоение миллиметрового диапазона и в системах мобильной сотовой связи, разрабатываемых комитетом по стандартизации 3GPP (Third Generation Partnership Project). Новая концепция построения мобильных сетей сотовой связи 5G, начиная со стандартов LTE-Release 15 и 5G New Radio (NR) [11], предполагает использование миллиметровых длин волн в диапазоне частот 24,25–52,6 ГГц. В частности, в этом диапазоне могут работать базовые станции малых сот с радиусом действия до 50–100 м, размещаемые в зонах покрытия существующих макросот в местах большого скопления пользователей. При этом передача большого объема данных между базовыми макро- и микростанциями будет осуществляться с использованием реконфигурируемой транспортной сети из небольших релейных ретрансляторов миллиметрового диапазона, обеспечивающих передачу данных со скоростью несколько гигабит в

секунду. Поэтому вопрос создания дешевых высокоскоростных систем радиосвязи на основе доступной элементной базы и с использованием различных высоконаправленных антенных систем [12–13] является весьма актуальным.

Основными недостатками приемо-передающих устройств, работающих в миллиметровом диапазоне, являются достаточно большие нелинейные искажения, вносимые усилителем мощности выходных каскадов передатчика, дисбаланс квадратурных (I/Q) компонент (далее – I/Q-дисбаланс) компонент и высокий уровень фазовых шумов. Эти недостатки общеизвестны и в основном связаны с технологией производства РЧ интегральных схем миллиметрового диапазона длин волн [14]. Проблема компенсации нелинейных искажений является особенно важной для приемо-передатчиков миллиметрового диапазона длин волн, использующих при производстве дешевые технологии.

Данный недостаток может быть устранен путем использования усилителей мощности в линейном режиме, что резко снижает их выходную мощность и коэффициент полезного действия. В ряде работ предложено уменьшать нелинейные искажения выходного сигнала путем применения специальных схем предискажения сигнала в радиопередатчике [15, 16] или дополнительных схем цифровой обработки сигнала в радиоприемнике [17], что существенно увеличивает сложность и стоимость устройства. Поэтому разработка новых способов борьбы с нелинейными искажениями является весьма актуальной, особенно для недорогих мобильных устройств.

В настоящей работе предложено и реализовано новое решение компенсации неизвестных нелинейных искажений передатчика и I/Q-дисбаланса с использованием статистической оценки этих искажений и адаптивного алгоритма демодуляции в радиоприемнике.

Для экспериментальной проверки эффективности предложенного решения был создан прототип приемо-передающего оборудования миллиметрового диапазона длин волн, использующий принцип программно-определяемого радио (SDR, *аббр. от анл.* Software Defined Radio) [18]. Оборудование предназначено для использования в небольших и дешевых релейных станциях, обеспечивающих решение актуальной технической задачи – беспроводной высокоскоростной передачи данных конечному пользователю или базовым станциям малых сот в гетерогенных системах мобильной связи 5G. Была решена задача разработки и программной реализации алгоритмов цифровой обработки сигналов на физическом уровне для приемо-передающего оборудования с программно-определяемым функционалом, разработана и реализована аппаратная часть, проведены экспериментальные измерения характеристик и полевые испытания прототипа.

1. Общее описание прототипа приемо-передающего оборудования диапазона 57–64 ГГц

Разработанный прототип состоял из двух приемо-передающих радиостанций, соединенных с управляющими персональными компьютерами (ПК). Общая структурная схема представлена на рисунке 1, из которой видно, что каждая из радиостанций представляла собой объединенный программно-аппаратный комплекс, состоящий из аналогового РЧ-блока, антенны (линзовой или рупорной), блока цифро-аналогового/аналого-цифрового преобразования (ЦАП/АЦП) и модуля цифровой обработки сигналов, реализованного на программируемой логической интегральной схеме (ПЛИС).

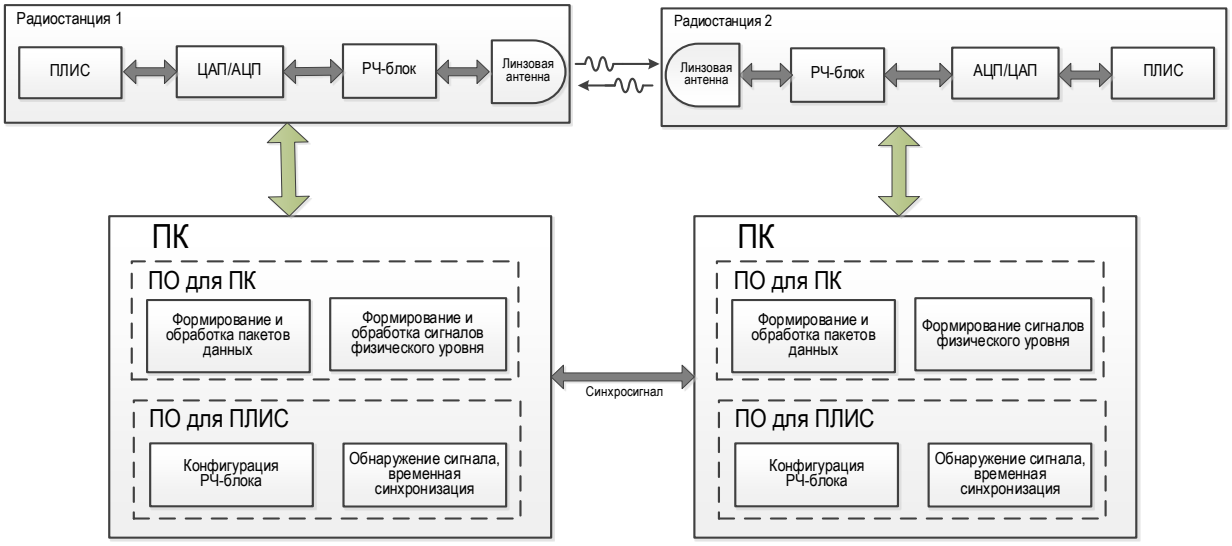


Рис. 1. Структурная схема прототипа приемопередающего оборудования диапазона частот 57–64 ГГц

Fig. 1. Block Diagram of the Transceiver Prototype of the Frequency Band 57–64 GHz

Специализированное программное обеспечение (ПО) прототипа состояло из ПО для ПК и ПО для ПЛИС. В состав первого входили блок формирования и обработки пакетов данных и блок формирования и обработки сигналов физического уровня. В состав второго – блок конфигурации РЧ-блока и блок обнаружения сигнала и временной синхронизации.

Прототип работал в режиме разделения приема и передачи по времени (TDD, *аббр. от англ.* Time Division Duplex). Выбранная структурная схема позволила программным образом реализовать различные сигнально-кодовые конструкции в радиопередатчике и соответствующие алгоритмы обработки этих сигналов в радиоприемнике, в модуле цифровой обработки сигналов. Конкретное построение отдельных блоков аппаратной и программной частей основывалось на имеющихся в свободном доступе РЧ-микросхемах ЦАП и АЦП и ПЛИС. При этом в модуле цифровой обработки сигналов радиоприемника возможно применение различных алгоритмов компенсации реальных нелинейных искажений, вносимых усилителем мощности в РЧ-блоке передатчика. Все ПО для прототипа было разработано с нуля и является оригинальным. За основу генерируемых сигнально-кодовых конструкций брался стандарт 802.11ad. Однако в силу технических ограничений РЧ-блока, максимальная полоса генерируемых сигналов составляла 800 МГц.

Общий вид одной приемо-передающей станции прототипа с управляющим персональным компьютером показан на рисунке 2, а основные характеристики приемо-передающей станции – в таблице 1.

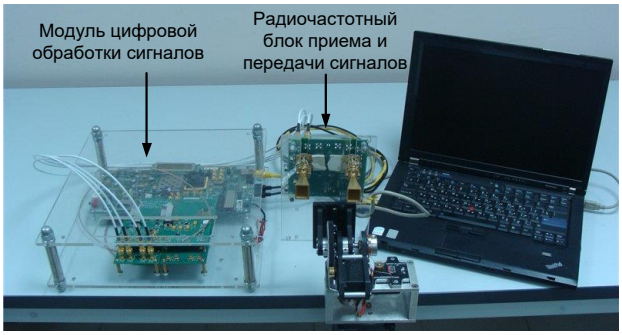


Рис. 2. Одна приемопередающая станция прототипа с управляющим персональным компьютером

Fig. 2. One Station of the Transceiver Prototype with a Personal Computer as a Controller

ТАБЛИЦА 1. Основные характеристики приемо-передающей станции

TABLE 1. Main Characteristics of the Receiving and Transmitting Station

Параметр	Значение
Несущая частота	57–64 ГГц
Эффективная полоса сигналов	до 800 МГц
Выходная мощность передатчика	до 10–12 дБм
Чувствительность приемника	–75 дБм

Ниже приведено краткое описание отдельных блоков и модулей приемо-передающей станции.

2. Аппаратная часть прототипа приемо-передающего оборудования

Аппаратная часть одной приемо-передающей станции включает в себя следующие основные блоки обработки сигналов.

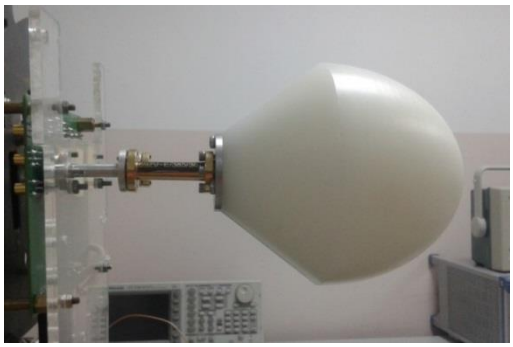
Антенны. Одна высоконаправленная антенна, работающая как на прием, так и на передачу сигналов, используется в полудуплексном режиме разделения передачи/приема по времени (TDD, *аббр. от англ.* Time Division Duplex). В прототипе также предусмотрена возможность использования двух высоконаправленных антенн (см. рисунок 2), работающих одна – на прием, а другая – на передачу сигналов при полном дуплексном режиме одновременной передачи/приема во времени с разделением сигналов по частоте (FDD, *аббр. от англ.* Frequency Division Duplex). Высоконаправленные антенны обеспечивают необходимое дополнительное пространственное разделение сигналов и их усиление в обоих направлениях без использования дуплексера. В зависимости от требуемой дальности работы линии связи и зоны покрытия в прототипе возможно применение как линзовых антенн с большим коэффициентом усиления до 31 дБи для обеспечения передачи данных на максимальное расстояние до 300 м, так и различных малогабаритных рупорных антенн с коэффициентом усиления от 12 до 21 дБи для обеспечения связи на более короткие расстояния. Основные характеристики используемых антенн приведены в таблице 2.

ТАБЛИЦА 2. Основные характеристики антенн миллиметрового диапазона длин волн, используемых в прототипе приемо-передающего оборудования

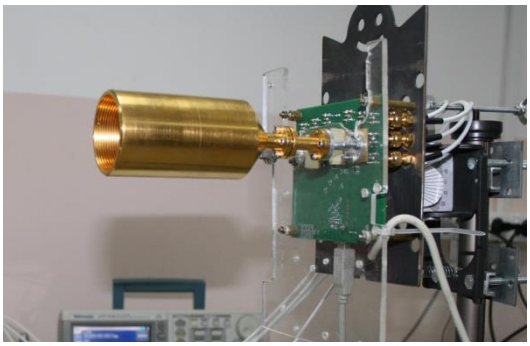
TABLE 2. The Main Characteristics of the Antennas of the Millimeter Wavelength Range Used in the Prototype of the Receiving and Transmitting Equipment

Наименование антенны	Коэффициент усиления, дБи	Ширина диаграммы направленности по уровню –3 дБ
Рупорная QSH-14125D0 Сечение круглое, d = 14 мм	12,3	30°
Рупорная 261E-20/387 Сечение прямоугольное 14×18 мм^2	19,8	18° (H plane) и 14° (V plane)
Рупорная QSH-14110D0 Сечение круглое, d = 38,5 мм	21	10°
Линзовая (d = 100 мм)	34,6	2,8°

Все типы используемых антенн имеют стандартный входной разъем, совместимый с прямоугольным волноводом WR15. Примеры линзовой и рупорной антенн в сборке с волноводами приведены на рисунке 3.



a)



b)

Рис. 3. Линзовая (a) и рупорная (b) антенны в сборке с волноводами

Fig. 3. Lens (a) and Horn (b) Antennas in Assembly with Waveguides

РЧ-блок приема и передачи сигналов выполняет усиление и фильтрацию аналоговых сигналов в диапазоне 57–64 ГГц, перенос их частоты с несущей в видеодиапазон (для радиоприемника) и обратно (для радиопередатчика), а также выделение квадратурных компонент сигнала.

Передающая часть РЧ-блока приема и передачи сигналов построена на базе микросхемы HMC6000 компании Hittite Microwave Corporation (см. URL: http://www.hittite.com/products/view.html/view/HMC6000). Микросхема HMC6000 является законченным решением передающего блока и уже имеет в своем составе следующие основные блоки: встроенный синтезатор частоты, режекторный фильтр, программируемый блок усиления промежуточной частоты, универсальный блок аналогового формирования ВЧ-сигнала. В результате синфазная и квадратурная составляющие сигнала, полученные с блока ЦАП модуля цифровой обработки сигналов, преобразуются в квадратурном модуляторе микросхемы HMC-6000 и на выходе получается ВЧ-сигнал в диапазоне 57–64 ГГц с заданными параметрами. Основные характеристики микросхемы HMC6000 приведены в таблице 3.

Приемная часть РЧ-блока приема и передачи сигналов построена на базе микросхемы HMC6001 компании Hittite Microwave Corporation (см. URL: http://www.hittite.com/products/view.html/view/HMC6001). Микросхема HMC6001 является законченным решением ВЧ-приемника и имеет в своем составе следующие основные блоки: малошумящий усилитель, режекторный фильтр, смеситель, фильтр промежуточной частоты, квадратурный демодулятор и встроенный синтезатор частоты. В результате на выходе микросхемы приемника получаются синфазная и квадратурная составляющие сигнала, которые передаются на вход блока АЦП через блок конфигурации и питания РЧ-блока. Опорная частота (285,714 МГц) для встроенного синтезатора частоты также подается с блока конфигурации и питания РЧ-блока (с тактового генератора). Усиление приемника может регулироваться в пределах: 2–67 дБ. Основные характеристики микросхемы HMC6001 приведены в таблице 3.

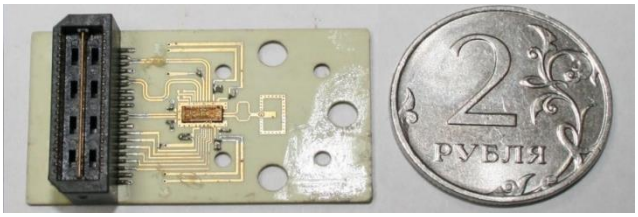
ТАБЛИЦА 3. Основные характеристики микросхем HMC6000 и HMC6001

TABLE 3. Main Characteristics of HMC6000 and HMC6001 Chips

TX HMC6000	Максимальная выходная мощность	12 дБм
	Максимальное усиление	38 дБ
	Диапазон перестройки усиления	17 дБ
	Шаг перестройки мощности	1.3 дБ
	Уровень шума	–32 дБм
RX HMC6001	Максимальное усиление	69 дБ
	Чувствительность	–75 дБм
	Диапазон частот	57–64 Ггц
	Коэффициент шума усилителя	7 дБ
	Рассеиваемая мощность	0,61 Вт
	Диапазон регулировки усиления	65 дБ

Микросхемы передатчика (HMC6000) и приемника (HMC6001) были смонтированы на отдельных разработанных в процессе выполнения работы печатных платах. Монтаж микросхем осуществлялся путем разварки. Общий вид блоков передатчика и приемника прототипа в сборке на платах показан на рисунке 4.

a)



b)

Рис. 4. Блоки передатчика (a) и приемника (b) прототипа в сборках на платах

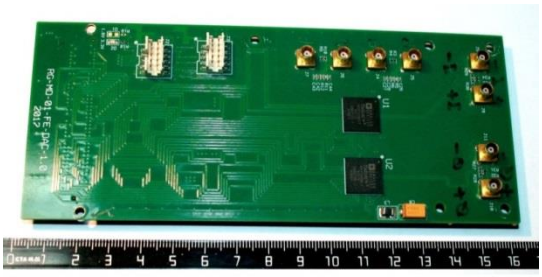
Fig. 4. Transmitter (a) and Receiver (b) Blocks of the Prototype in the Assemblies on the Boards

Управление блоками передатчика и приемника (микросхемами HMC6000 и HMC6001) осуществлялось непосредственно с ПК через блок конфигурации и питания РЧ-блока приема и передачи сигналов. Все необходимые напряжения питания формировались в блоке конфигурации и питания радиомодуля.

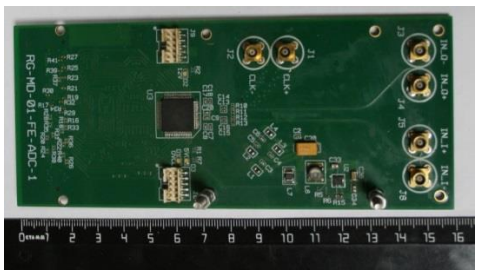
Таким образом, РЧ-блок приема и передачи сигналов сопрягает используемую антенную систему и блоки обработки сигналов на видеочастоте.

Блоки ЦАП и АЦП осуществляют сопряжение аналогового РЧ-блока приема и передачи сигналов с блоком цифровой обработки сигналов.

Блок ЦАП построен на базе двух микросхем AD9734 (см. URL: http://www.analog.com/static/imported-files/data_sheets/AD9734_9735_9736.pdf) с разрядностью квантования сигналов 10 бит каждая. Использование двух микросхем необходимо для одновременного ЦАП синфазной и квадратурной компонент ВЧ-сигнала. Каждый канал данного блока ЦАП позволяет осуществлять ЦАП сигналов с шириной полосы до 600 МГц, что позволяет формировать РЧ-сигналы с полосой до 1,2 ГГц. Для передачи высокоскоростных сигналов (данных и тактовых сигналов) от блока цифровой обработки сигналов к блоку ЦАП используются согласованные дифференциальные линии передачи с волновым сопротивлением 100 Ом. Аналоговые выходы синфазного и квадратурного каналов блока ЦАП подключаются через коаксиальные 50-омные кабели и аттенюаторы к РЧ-разъемам блока конфигурации и питания радио для дальнейшего соединения с блоком передатчика, что позволяет формировать РЧ-сигналы с полосой до 1,2 ГГц. Общий вид платы блока ЦАП показан на рисунке 5a.



a)



b)

Рис. 5. Общий вид блоков ЦАП (a) и АЦП (b)

Fig. 5. General View of DAC (a) and ADC (b) Blocks

В блоке высокоскоростного АЦП используется одна микросхема двухканального АЦП компании Maxim Integrated MAX105 (см. URL: http://datasheets.maximintegrated.com/en/ds/MAX105.pdf). Она может преобразовывать сигналы с шириной полосы до 400 МГц в каждом квадратурном канале, что позволяет обрабатывать РЧ-сигналы с полосой до 800 МГц. Микросхема АЦП MAX105 имеет встроенный последовательно-параллельный преобразователь, что позволяет снизить скорость принимаемых модулем ПЛИС цифровых данных в два раза за счет параллельной передачи двух битовых потоков для каждого разряда в каждом канале. Соединение с блоком цифровой обработки сигналов выполнено аналогично блоку ЦАП, с использованием согласованной низковольтной дифференциальной передачи сигналов (LVDS, *аббр. от англ.* Low Voltage Differential Signaling,) шины для передачи данных и тактовых сигналов синхронизации. Аналоговые входы синфазного и квадратурного каналов блока АЦП подключаются через коаксиальные 50-омные кабели к РЧ-разъемам блока конфигурации и питания радио для дальнейшего соединения с блоком приемника. Общий вид платы блока АЦП показан на рисунке 5b.

Тактовая частота с блока генерации тактового сигнала подается в блоки ЦАП и АЦП, управление сигналами и конфигурация которыми осуществляется с ПЛИС платы блока цифровой обработки сигналов.

Блок цифровой обработки сигналов и управления для каждой приемо-передающей станции построен на базе отладочной платы KC705 производства компании Xilinx (см. http://www.xilinx.com/support/documentation/boards_and_kits/kc705/ug810_KC705_Eval_Bd.pdf).

Основным элементом отладочной платы KC705 является программируемая логическая интегральная схема серии Kintex7 XC7K325T-2FFG900C производства компании Xilinx (см. URL: http://www.xilinx.com/support/documentation/data_sheets/ds180_7Series_Overview.pdf). Выбор отладочной платы с этой микросхемой обусловлен наличием в ее составе необходимых аппаратных ядер для реализации высокоскоростного приема, генерации и обработки сигналов и данных. Кроме того, на плате присутствуют используемые в ходе работы с экспериментальной установкой Ethernet блок, USB – блок и собственный блок питания. Основные характеристики ПЛИС Xilinx XC7K325T-2FFG900C приведены в таблице 4.

Вспомогательные блоки. В передающей и приемной частях РЧ-блока имелись также идентичные вспомогательные блоки: генерации тактового сигнала для формирования ВЧ тактовых сигналов для АЦП и ЦАП; конфигурации и питания РЧ-блока приема и передачи сигналов. Кроме того, на пла-

тах передатчика и приемника были смонтированы специально разработанные волноводно-микрополосковые переходы для подключения различных антенн через волноводный интерфейс типа WR15.

ТАБЛИЦА 4. Основные характеристики ПЛИС Xilinx XC7K325T-2FFG900C

TABLE 4. Main Characteristics of FPGA Xilinx XC7K325T-2FFG900C

Параметр		Количество
Число логических ячеек		326080
Число вводов-выводов	Общее число	500
	Число дифференциальных пар	240
Встроенные аппаратные блоки	Блоки цифровой обработки	840
	Блоки PCI Express	1
	Высокоскоростные приемопередатчики GTX	16

Персональный компьютер входит в состав каждой приемо-передающей станции созданного прототипа и имеет в рамках данного проекта двойное назначение. С одной стороны, ПК используется как источник и приемник потока данных, передаваемых через разрабатываемый прототип, т. е. как клиентское устройство. С другой стороны, на ПК производится выполнение части алгоритмов цифровой обработки сигналов и измерение характеристик работы прототипа в целом.

3. Программная часть прототипа приемо-передающего оборудования

Программная часть разработанного прототипа содержит процедуры конфигурации блоков аппаратной части, реализует алгоритмы цифровой обработки сигналов в радиопередатчике и приемнике, а также формирует пакеты данных в соответствии с выбранным форматом сигналов физического уровня. К программной части разрабатываемого решения относятся как программные прошивки для ПЛИС, так и ПО, предназначенное для выполнения на ПК (см. рисунок 1).

ПО для ПЛИС конфигурирует микросхему Kintex7 для выполнения задач управления РЧ-блоком приема и передачи сигналов и первичной высокоскоростной цифровой обработки сигналов, поступающих с АЦП (в радиоприемнике) или предназначенных для подачи на ЦАП (в радиопередатчике). ПО для ПЛИС реализовано с использованием языка описания аппаратуры (HDL, *аббр. от англ.* Hardware Description Language) Verilog. Управление РЧ-блоком приема и передачи сигналов заключается в формировании управляющих сигналов и команд, посылаемых на предусмотренные в РЧ-блоке приема и передачи сигналов выводы и интерфейсы. Функционал цифровой обработки сигналов в ПЛИС заключается, во-первых, в буферизации потока отсчетов сигнала во временно́й области для того, чтобы обеспечить строго периодическое поступление этих отсчетов на ЦАП в радиопередатчике и строго периодический их съем с АЦП в радиоприемнике. Во-вторых, в ПЛИС были реализованы алгоритмы первичного обнаружения пакета и временной синхронизации работы приемника с принимаемыми сигналами.

ПО на ПК в режиме передачи обеспечивает в блоке формирования и обработки пакетов данных преобразование битового потока, поступающего от верхних сетевых уровней, в соответствии с протоколом выбранного стандарта физического уровня. В рамках этой задачи выполняются такие процедуры, как разбиение битового потока на пакеты заданной длины, генерация заголовков физического уровня и контрольных сумм, а также объединение этих информационных элементов в законченные пакеты, готовые для отправки в блок формирования и обработки сигналов физического уровня. В режиме приема происходит обратное преобразование сигналов. Блок формирования и обработки сигналов физического уровня, в свою очередь, выполняет процедуры преобразования пакетов данных во временные отсчеты сигналов в соответствии с форматом сигналов выбранного стандарта физического уровня. Данный блок реализует следующие процедуры: помехоустойчивое кодирование/декодирование битового потока на основе LDPC-кодов (*от англ.* Low-Density Parity Checking Codes – коды с малой плотностью проверки на четность) модулятор/демодулятор, взаимнооднозначно преобразующий информационную битовую последовательность в цифровые отсчеты сигнала при передаче и обратно при приеме в соответствии с заданной схемой модуляции: 2-ФМ (BPSK), 4-ФМ (QPSK), 16-КАМ (16-QAM).

Для достижения высокого уровня универсальности ПО на ПК реализовано на высокоуровневых языках программирования (C, C++, Matlab).

4. Функциональная схема приемо-передающего оборудования

Общая функциональная схема разработанного приемо-передающего оборудования представлена на рисунке 6.

На стороне передатчика в программном блоке передатчика в реализованных программных модулях осуществляется следующий набор основных операций: помехоустойчивое кодирование (модуль кодера), модуляция сигналов (модуль модулятора), формирование пакета физического уровня (модуль формирования пакета) и формирование выходных сигналов в заданной полосе (модуль формирования выходного сигнала). Модуль кодера осуществляет преобразование информационной последовательности бит и вносит в нее избыточность, необходи-

мую для исправления ошибок передачи на приемной стороне помехоустойчивым декодером. Результат работы кодера поступает на вход модулятора, преобразующего битовую последовательность в отсчеты комплексной амплитуды сигнала, предназначенного для передачи. Модуль формирования пакета данных отвечает за структуру пакета: добавляет служебные сигналы, необходимые для детектирования пакета, и пилотные сигналы для оценки канала и его эквализации. Сформированная последовательность отсчетов сигналов в пакете подвергается процедуре цифровой фильтрации с целью обеспечения требуемой полосы передачи сигнала.

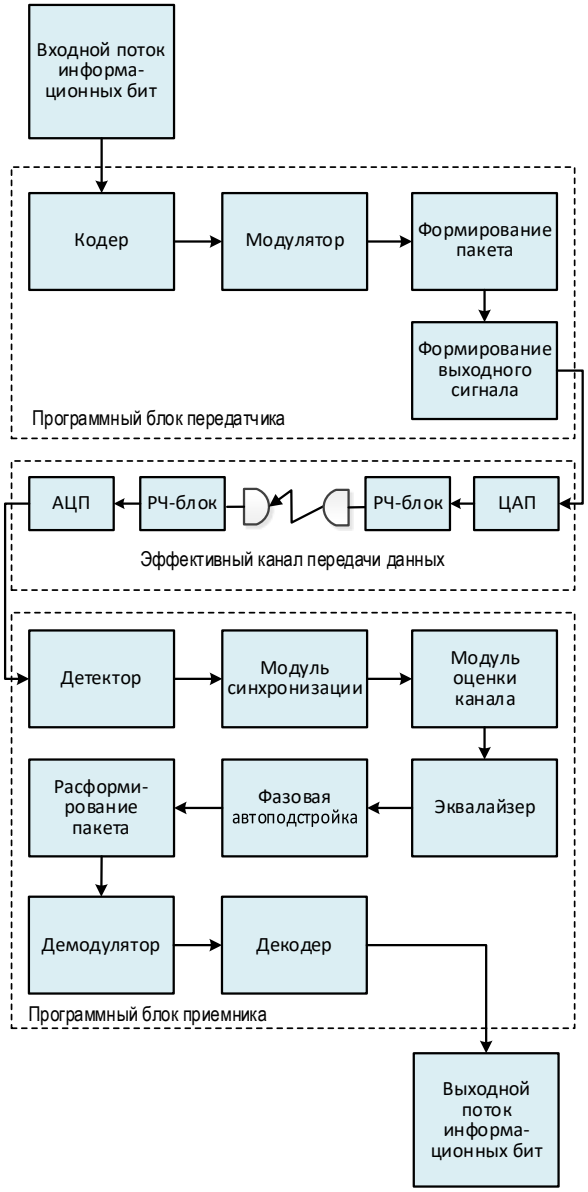


Рис. 6. Общая функциональная схема разработанного приемо-передающего оборудования.

Fig. 6. Functional Diagram of the Developed Transceiver Equipment

Эффективным каналом передачи для цифрового сигнала является участок функциональной схемы от входа ЦАП до выхода АЦП, включающий в себя РЧ-части и антенные системы приемной и передающей аппаратуры, а также среду распространения сигнала. В процессе передачи сигнал подвергается искажениям, вызванным частотной селективностью канала, временно́й и частотной расстройкой гетеродинов приемной и передающей аппаратуры.

На приемной стороне в программном блоке приемника модуль детектора осуществляет корреляционную обработку служебных сигналов в преамбуле пакета и выносит решение об обнаружении передаваемого пакета на фоне шума. В модуле синхронизации устраняется частотно-временная расстройка между опорными сигналами приемника и передатчика. Для восстановления искаженной каналом формы сигнала проводится оценка передаточной характеристики канала связи в модуле оценки канала. Восстановленные в программном модуле эквалайзера сигналы подвергаются более точной фазовой автоподстройке, позволяющей компенсировать набеги фазы сигнала, вызванные флуктуациями несущей частоты в аппаратуре приемника и передатчика. Модуль расформирования пакета отвечает за разделение служебных и информационных сигналов в принятом пакете. Информационные сигналы подвергаются демодуляции в соответствующем модуле, и полученная кодовая последовательность поступает на вход декодера, отвечающего за исправление ошибок при передаче.

Для представленных на функциональной схеме (см. рисунок 6) модулей цифровой обработки сигналов в ходе создания прототипа был проведен отбор эффективных алгоритмов, осуществлена их программная реализация и тестирование.

5. Формат сигналов физического уровня

Основой для определения формата сигналов физического уровня в прототипе является принятый стандарт Wi-Fi миллиметрового диапазона IEEE 802.11ad с определенными модификациями, учитывающими особенности реализации и предполагаемого использования разрабатываемого приемо-передающего оборудования. В прототипе используется режим передачи с модуляцией на одной несущей частоте (Single Carrier PHY).

Структура используемых пакетов данных повторяет структуру, определенную в стандарте IEEE 802.11ad [7], за исключением специального поля (BFT, *аббр. от англ.* Beamforming Training Field), встраиваемого в конце пакета и предназначенного стандартом для подстройки диаграмм направленности антенных систем в мобильных приложениях. При формировании структуры пакета в рамках настоящей работы были учтены также особенности аппаратной части прототипа приемо-передающего оборудования, что привело к изменению парамет-

ров сигналов по сравнению со стандартом IEEE 802.11ad. Частота дискретизации сигналов для пакета физического уровня составляет 800 МГц, а ограничение на объем оперативной памяти составляет 46 872 отсчета.

Разработанная в рамках настоящей работы структура пакета физического уровня представлена на рисунке 7.

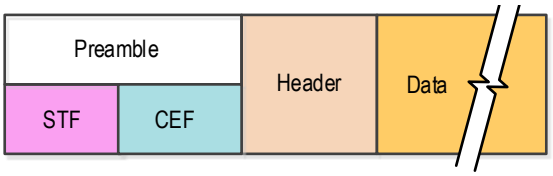


Рис. 7. Общая структура пакета физического уровня

Fig. 7. General Structure of One Physical Layer Package

Пакет физического уровня состоит из преамбулы (Preamble), заголовка пакета (Header) и поля передачи полезной информации (Data). Preamble состоит из двух частей: короткой тренинговой последовательности (STF, *аббр. от англ.* Short Training Field) и длинной – для оценки канала (CEF, *аббр. от англ.* Channel Estimation Field). Поле STF используется для первоначального обнаружения пакета и установления временной и частотной синхронизации, поле CEF – для оценки импульсной переходной характеристики канала связи, применяемой в дальнейшем для процедуры эквализации канала, предшествующей демодуляции принятых сигналов. Обе тренинговые последовательности состоят из набора идущих подряд последовательностей Голея и имеют структуру соответствующих тренинговых последовательностей стандарта IEEE 802.11ad.

Header содержит служебную информацию о длине пакета, схеме модуляции и кодирования (MCS, *аббр. от англ.* Modulation and Coding Scheme), используемой в поле Data и инициализирующую последовательность скремблера. Индекс MCS имеет длину 8 бит и содержит информацию о полосе, занимаемой сигналом, скорости кодирования и типе модуляции, используемыми при передаче поля данных. Последний бит 8-битового слова является резервным. Исходное 8-битовое слово модулируется с помощью дифференциальной модуляции π/2–DBPSK [7] и интерполируется в 8 раз. Таким образом, заголовок имеет длину 64 отсчета и передается в узкой полосе 100 МГц. Данная структура заголовка существенно повышает надежность его приема из-за увеличения отношения сигнал/шум на входе радиоприемника. Кроме того, дифференциальная модуляция позволяет осуществить некогерентный прием заголовка, что дает возможность закончить за это время точную оценку характеристик канала и эквалайзера и подготовиться к приему информационных символов в поле данных в синхронном режиме.

Приемо-передающее оборудование поддерживает 12 MCS, предусмотренных стандартом IEEE 802.11ad для передачи данных. При этом используются три вида модуляции: 2-ФМ (π/2-BPSK), 4-ФМ, 16-КАМ и четыре скорости кодирования: 1/2, 3/4, 5/8 и 13/16. В качестве схемы помехоустойчивого кодирования используются стандартные коды с малой плотностью проверок на четность с длиной кодового слова 672 бита. Список используемых MCS с указанием достигаемых скоростей передачи на уровне одного пакета физического уровня без учета затрат времени на передачу служебных символов и соответствующие скорости передачи данных с учетом потерь на служебные символы при полосе сигналов 800 МГц приведен в таблице 5.

ТАБЛИЦА 5. Скорость передачи данных на физическом уровне для различных схем модуляции и кодирования

TABLE 5. Data Transfer Rate at the Physical Level for Various Modulation and Coding Schemes

MCS-индекс	Тип модуляции	Скорость кода	SINR, дБ	Скорость передачи данных, Мбит/с	
				без учета служебных символов	с учетом служебных символов в пакете
1110000	π/2-BPSK	1/2	1,2	400	334
1111000	π/2-BPSK	5/8	2,2	500	418
1110100	π/2-BPSK	3/4	3,0	600	501
1111100	π/2-BPSK	13/16	3,7	650	543
1110001	QPSK	1/2	4,2	800	669
1111001	QPSK	5/8	5,2	1000	836
1110101	QPSK	3/4	6,0	1200	1003
1111101	QPSK	13/16	6,7	1300	1086
1110011	16QAM	1/2	10,2	1600	1337
1111011	16QAM	5/8	11,4	2000	1671
1110111	16QAM	3/4	12,5	2400	2006
1111111	16QAM	13/16	13,2	2600	2173

Примечание: SINR – отношения сигнал/помеха+шум

В силу того, что разрабатываемое приемо-передающее оборудование предназначено для передачи информации в конфигурации «точка-точка» и равноправности направлений связи, в прототипе использован режим передачи с временным разделением восходящей и нисходящей линий связи с адаптивным разделением времени в зависимости от объема передаваемых данных в ту и в другую сторону. Для обеспечения возможности передачи данных на различные расстояния с максимальной пропускной способностью разработанное приемо-передающее оборудование поддерживает дискретное изменение полосы спектра передаваемых сигналов (100, 200, 400 и 800 МГц). Использование принципа дискретного масштабирования рабочей полосы частот в разработанном

SDR-модеме и использование набора оригинальных линзовых и рупорных антенн позволяет в перспективе создать линейку приемо-передающего оборудования миллиметрового диапазона длин волн с возможностью гибкого выбора необходимой скорости передачи информации, дальности связи и стоимости оборудования.

6. Испытания и экспериментальные измерения характеристик прототипа

6.1. Методика испытаний разработанного оборудования

С целью формирования и последующей обработки принятых пакетов физического уровня, которые передаются и принимаются аппаратной частью приемо-передающего оборудования, используется специальное ПО, реализованное в среде Matlab на ПК. Разработанное ПО используется для проведения испытаний и измерения характеристик прототипа приемо-передающего оборудования. ПО позволяет генерировать пакеты в соответствии с выбранным форматом сигналов физического уровня (вида модуляции, скорости кодирования, полосы передачи данных и т. п.), распаковывать принятые пакеты, вычислять вероятности битовых (BER) и пакетных (BLER) ошибок, оценивать суммарное SINR на входе радиоприемника, вычислять и выводить на экран дополнительную информацию с помощью удобного графического интерфейса пользователя.

Для быстрой оценки качества приема сигналов в процессе проведения экспериментов для каждого пакета физического уровня (серии пакетов) ПО позволяет осуществлять визуализацию квадратурных компонент принимаемых сигналов на выходе демодулятора приемника перед декодером. Это осуществляется путем построения модуляционных «созвездий» (Constellations) на плоскости квадратурных компонент и расчета величины среднего квадратичного отклонения (EVM, *аббр. от англ.* Error Vector Magnitude) принимаемых сигналов от их идеальных (невозмущенных) значений. Величина EVM, измеренная в дБ, оперативно используется для оценки отношения мощности полезного сигнала к суммарной мощности всех помех: теплового шума, фазового шума и межсимвольных помех, связанных с многолучевостью.

Разработанное специальное ПО дает возможность также проводить в реальном времени измерения импульсных характеристик реального канала распространения сигнала с разрешающей способностью 1.25 нс (эквивалентное разрешение по расстоянию составляет 38 см). Это позволяет выявлять причины изменения характеристик системы связи, вызванных переотражениями сигнала от местных предметов (многолучевостью канала), контролировать работу блока эквалайзера в радиоприемнике и отдельно оценивать уровень и влияние межсимвольных помех на работу приемо-передающего оборудования.

Измерения вероятностей битовых ошибок проводились не менее чем по 20 пакетам физического уровня, каждый из которых имел поле данных длиной 43 136 символов, из которых 37 632 символа использовались для передачи закодированных информационных сообщений. В таблице 6 приведено количество LDPC кодовых блоков и количество передаваемых бит в одном пакете физического уровня для различных видов модуляций и скоростей кодирования.

ТАБЛИЦА 6. Количество кодовых блоков и бит в составе одного пакета для полосы 800 МГц

TABLE 6. Number of Blocks and Bits in One Packet for the 800 MHz Band

Вид модуляции	Число кодовых блоков	Общее число передаваемых бит	Число информационных бит			
			1/2	5/8	3/4	13/16
BPSK	56	37632	18816	23520	28224	30576
QPSK	112	75264	37632	47040	56448	61152
16QAM	225	151200	75600	94500	113400	122850

Несложно показать, что при полном отсутствии битовых ошибок в серии измерений BER по 20 пакетам для самой низкой спектральной эффективности (BPSK, скорость кодирования 1/2) с доверительной вероятностью $\beta = 0{,}95$ уровень битовых ошибок не превысит значение 10^{-5}. При этом с такой же доверительной вероятностью уровень пакетных ошибок (PER, *аббр. от англ.* Packet Error Ratio) не превысит значение 0,0025. Аналогичным образом при проведении экспериментальных измерений оценивались верхние границы доверительных интервалов для вероятностей битовых и пакетных ошибок и для других модуляций и скоростей кодирования.

Тестирование и экспериментальное измерение характеристик разработанного прототипа приемо-передающего оборудования проводилось в различных режимах работы: как в лабораторных условиях, так и в близких к реальным условиям работы. В настоящую статью включены экспериментальные результаты для режимов работы с сигналами с полосой 800 МГц, при которых возможно достижение максимальных пропускных способностей. Проведенные испытания работы прототипа с сигналами с меньшей полосой (100, 200, 400 МГц) и, соответственно, меньшей скоростью передачи данных, показали, что при таких режимах работы обеспечивается бóльшая дальность радиосвязи, поскольку при этом достигаются бóльшие отношения сигнал/шум и эти сигналы существенно меньше подвержены межсимвольной интерференции.

6.2. Лабораторные исследования характеристик прототипа

В ходе лабораторных экспериментальных исследований прототипа приемо-передающего оборудования были проведены две серии измерений: в обычной лабораторной комнате и в специальной безэховой камере. При этих исследованиях измерялись вероятности битовых и пакетных ошибок и уровни SINR для разных модуляций и скоростей кодирования. Для контроля работы системы связи также выводились диаграммы созвездий для различных модуляций, вероятности битовых ошибок без использования схемы помехоустойчивого кодирования и измеренные импульсные переходные характеристики канала распространения сигнала.

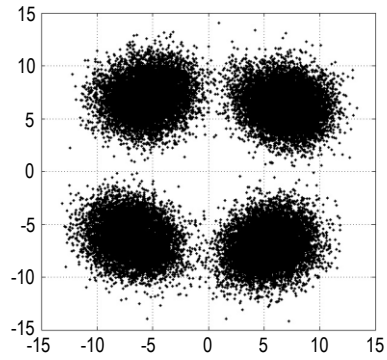
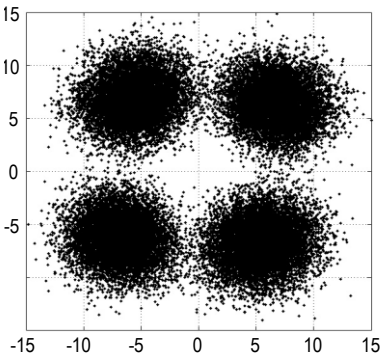
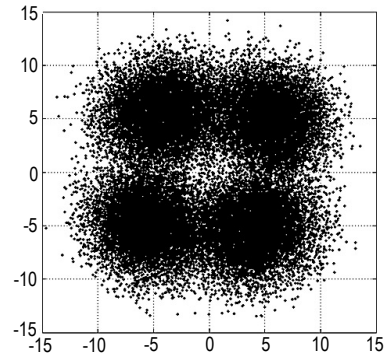
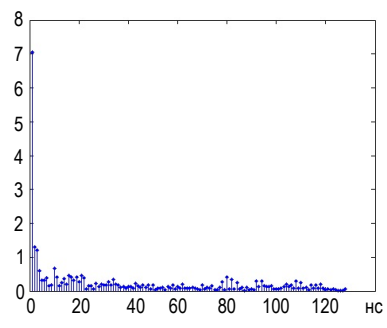
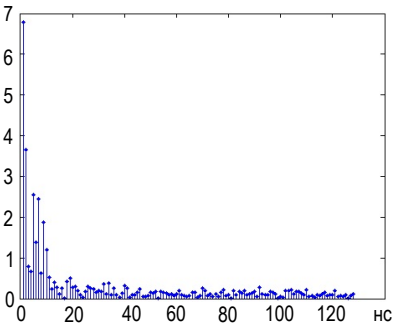
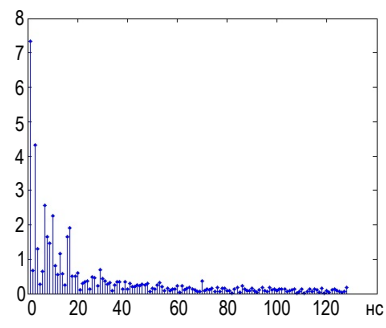
6.2.1. Измерения в лабораторной комнате

При измерениях в лабораторной комнате в качестве излучающей и приемной антенн использовался открытый конец волновода типа WR15 (конфигурация антенн «волновод-волновод»). Устанавливалась заниженная мощность передатчика P_{TX} = 4,2 дБм одинаковая для всех модуляций и скоростей кодирования. Расстояние между передатчиком и приемником (L) варьировалось: 1; 1,5; 2 м.

В таблице 7 представлены примеры диаграмм рассеяния сигналов («созвездия») с полосой 800 МГц для QPSK-модуляции. Под диаграммами рассеяния сигналов приведены: измеренные величины битовых ошибок без использования помехоустойчивого кодирования (BER), битовые и пакетные ошибки с использованием помехоустойчивого кодирования (BER-LDPC и PER) для скорости кодирования 3/4, уровни SINR и пропускные способности (TH) на уровне одного пакета физического уровня с учетом затрат на служебные символы. Измерения характеристик проводились в лабораторной комнате путем усреднения по 20 пакетам физического уровня и показали, что при конфигурации антенн типа «волновод-волновод» информационные пакеты принимаются с малой вероятностью ошибок (< 10^{-2}) только на малых расстояниях до 1,5–2 м. На расстояние 2 м устойчиво передавались пакеты только с BPSK модуляцией. Пакеты с QPSK модуляцией передаются на это расстояние уже с недопустимо большой вероятностью ошибки (PER = 0,87).

ТАБЛИЦА 7. Диаграммы рассеяния сигналов («созвездия») для QPSK модуляций и импульсные переходные характеристики канала связи при измерениях в лабораторной комнате

TABLE 7. Signal Scattering Diagrams ("Constellations") for QPSK Modulations and Pulse Transient Characteristics of the Communication Channel during Measurements in the Laboratory Room

L = 1 м	L = 1,5 м	L = 2 м
		
BER = 1,0 × 10^{-3} BER-LDPC < 2,6 × 10^{-6} PER < 1,3 × 10^{-3} SINR = 11,6 дБ TH = 1003 Мбит/сек	BER = 2,7 × 10^{-3} BER-LDPC < 2,6 × 10^{-6} PER < 1,3 × 10^{-3} SINR = 7,0 дБ TH = 1003 Мбит/сек	BER = 3,0 × 10^{-2} BER-LDPC =2,6 × 10^{-2} PER = 0,87 SINR = 5,1 дБ TH – срыв передачи
Импульсные переходные характеристики канала		
		

Из измеренных импульсных переходных характеристик канала видно, что наблюдается сильная многолучевость, обусловленная лучами, отраженными от стен и различных предметов, находящихся в комнате, которые из-за широких диаграмм направленности открытых волноводов составляют существенную часть мощности принимаемых сигналов, т. е. канал является существенно частотно-селективным. При этом показательное время спадания импульсной переходной характеристики равно 20 нс. Это в два раза превышает длительность используемого защитного интервала, равного 10 нс, что приводит к сильной межсимвольной интерференции. Полученные экспериментальные результаты соответствуют расчетам бюджета линии связи и еще раз демонстрируют, что для систем миллиметрового диапазона длин волн необходимо использовать высоконаправленные антенны с большими коэффициентами усиления.

На диаграммах рассеяния QPSK при больших SINR (L = 1 м) отчетливо виден вклад фазовых шумов передающей и приемной частей РЧ-блока приема и передачи сигналов, и/или нелинейности усилителя передающей части, приводящие к размытию диаграмм рассеяния по окружности. Видно, что вклад этих искажений сигнала может стать определяющим фактором для передачи сигналов с модуляциями высокого порядка при больших SINR.

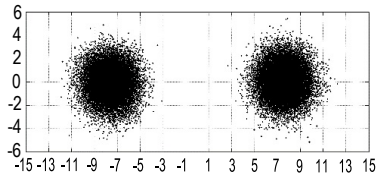
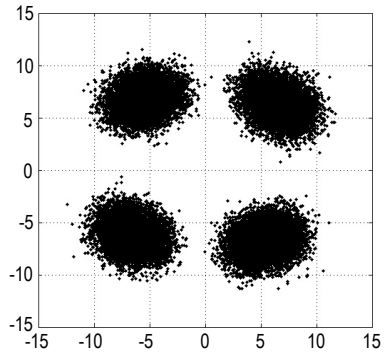
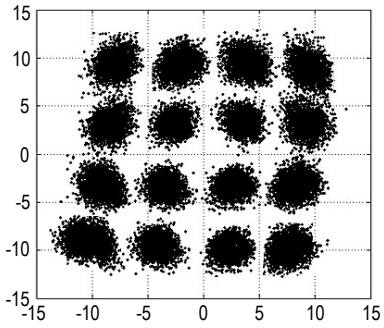
6.2.2. Измерения в безэховой камере

При измерениях в безэховой камере в радиопередатчике в качестве излучающего элемента служит прямоугольная рупорная антенна (см. таблицу 1), а на приемном конце линии связи использовался открытый конец волновода типа WR15 (конфигурация антенн «рупор – волновод»). Расстояние между передатчиком и приемником является фиксированным 5 м и определяется максимальным размером безэховой камеры. Установленная заниженная мощность передатчика P_{TX} = = 4,2 дБм одинакова для всех модуляций и скоростей кодирования.

В таблице 8 представлены примеры диаграмм рассеяния сигналов («созвездия») с полосой 800 МГц для BPSK, QPSK и 16-QAM модуляций. Под диаграммами рассеяния сигналов приведены данные аналогичные таблице 7. Измерения характеристик проводились путем усреднения по 20 пакетам физического уровня.

ТАБЛИЦА 8. Диаграммы рассеяния сигналов («созвездия») для BPSK, QPSK и 16-QAM модуляций при измерениях в безэховой камере

TABLE 8. Signal Scattering Diagrams ("Constellations") for BPSK, QPSK and 16-QAM Modulations during Measurements in an Anechoic Chamber

BPSK	QPSK	16-QAM
		
BER < 4,0 × 10^{-6} BER-LDPC < 5,3 × 10^{-6} PER < 2,7 × 10^{-3} SNR = 14,8 дБ TH = 501 Мбит/сек	BER = 2,8 × 10^{-4} BER-LDPC < 2,6 × 10^{-6} PER < 1,3 × 10^{-3} SNR = 14,8 дБ TH = 1003 Мбит/сек	BER = 5,8 × 10^{-3} BER-LDPC = 6,7*10^{-5} PER = 3,3 × 10^{-2} SNR = 14,8 дБ TH = 2006 Мбит/сек

Исследования в безэховой камере показали, что канал связи в безэховой камере близок к идеальному однолучевому каналу с аддитивным белым гауссовским шумом. В этих условиях при конфигурации антенн типа «рупор – волновод» на расстоянии 5 м величина SINR достигала значения в 14,5 дБ и информационные пакеты для всех используемых видов модуляций при скорости кодирования 3/4 принимались с малой вероятностью пакетных ошибок (PER < 0,05). В то же время на диаграммах рассеяния видны относительно сильные искажения созвездий, связанные с I/Q-дисбалансом, фазовыми шумами и нелинейностью передатчика. Без применения специальных схем компенсации эти искажения становятся основным ограничивающим фактором для модуляции 16-QAM.

6.3. Полевые испытания

При полевых испытаниях проводились исследования характеристик прототипа приемо-передающего оборудования для базового расстоя-

ния между передатчиком и приемником 100 м для нескольких сценариев окружающей обстановки. Однако при всех измерениях приемная и передающая станции находились на линии прямой видимости (сценарий LOS, *аббр. от англ.* Line-of-Sight). Измерялись битовые и пакетные ошибки с использованием помехоустойчивого кодирования (BER-LDPC и PER) для для всех 12 MCS. Использовалась антенная конфигурация «линза – рупор» при мощности передатчика 4,2 дБм.

Проведенные измерения показали, что для всех сценариев окружающей обстановки из-за высокой направленности используемых антенн импульсная переходная характеристика канала определялась в основном одним прямым лучом и результаты измерений повторялись для всех сценариев расположения антенн в пределах погрешностей. В таблице 9 приведены усредненные результаты измерения BER-LDPC и PER по 10 сериям измерений. Каждая серия состояла из 20 пакетов физического уровня. В таблице также приведены соответствующие скорости передачи на уровне одного пакета физического уровня без учета затрат времени на передачу служебных символов и соответствующие скорости передачи данных с учетом потерь на служебные символы.

ТАБЛИЦА 9. Сводная таблица результатов экспериментальных измерений при полевых испытаниях L = 100 м

TABLE 9. Summary Table of the Results of Experimental Measurements during Field Tests L = 100 m

Тип модуляции	Скорость кода	Скорость передачи, Мбит/с		BER -LDPC	PER
		на физическом уровне	в пакете		
π/2-BPSK	1/2	400	334	$<8{,}0 \times 10^{-6}$	$<2{,}7 \times 10^{-3}$
π/2-BPSK	5/8	500	418	$<6{,}4 \times 10^{-6}$	$<2{,}7 \times 10^{-3}$
π/2-BPSK	3/4	600	501	$<5{,}3 \times 10^{-6}$	$<2{,}7 \times 10^{-3}$
π/2-BPSK	13/16	650	543	$<4{,}9 \times 10^{-6}$	$<2{,}7 \times 10^{-3}$
QPSK	1/2	800	669	$<4{,}0 \times 10^{-6}$	$<1{,}3 \times 10^{-3}$
QPSK	5/8	1000	836	$<3{,}2 \times 10^{-6}$	$<1{,}3 \times 10^{-3}$
QPSK	3/4	1200	1003	$<2{,}6 \times 10^{-6}$	$<1{,}3 \times 10^{-3}$
QPSK	13/16	1300	1086	$<2{,}4 \times 10^{-6}$	$<1{,}3 \times 10^{-3}$
16-QAM	1/2	1600	1337	$<2{,}0 \times 10^{-6}$	$<6{,}7 \times 10^{-4}$
16-QAM	5/8	2000	1671	$2{,}2 \times 10^{-5}$	$9{,}2 \times 10^{-3}$
16-QAM	3/4	2400	2006	$8{,}4 \times 10^{-5}$	$4{,}2 \times 10^{-2}$
16-QAM	13/16	2600	2173	$2{,}8 \times 10^{-4}$	$1{,}2 \times 10^{-1}$

Полученные результаты показывают, что разработанный прототип обеспечивает скорость передачи данных в пакете до 2 Гбит/с на расстоянии 100 м при использовании 16-QAM модуляции со скоростью кода 3/4 при допустимых пакетных ошибках PER < 0,05. Однако при скорости кодирования 13/16 вероятность пакетных ошибок достигает 12 %. Следует отметить, что увеличение мощности передатчика при модуляции 16-QAM не позволило снизить вероятности пакетных ошибок, по всей видимости, из-за возникающих нелинейных искажений сигналов в радиопередатчике на фоне неустранимого остаточного фазового шума.

7. Применение адаптивного алгоритма демодуляции для компенсации I/Q-дисбаланса и нелинейных искажений сигнала на стороне передатчика

Проведенные экспериментальные исследования выявили, что используемые в РЧ-блоке прототипа микросхемы HMC6000 и HMC6001 компании Hittite Microwave Corporation обладают высоким уровнем фазовых шумов, I/Q-дисбалансом около 8–10 % и большими нелинейными искажениями при максимальных мощностях (10–12 дБм). Следует отметить, что присутствие сильных фазовых шумов типично для систем связи миллиметрового диапазона длин волн [19, 20]. Для борьбы с фазовыми шумами в прототипе использовался известный алгоритм, компенсирующий линейный тренд фазовых набегов в радиоприемнике на основе периодических коротких пилотных сигналов [7], подробно описанный в работе [21].

Для компенсации I/Q-дисбаланса модулятора/демодулятора и нелинейных искажений сигнала, вызванных неидеальной работой усилителя мощности в радиопередатчике, был реализован адаптивный алгоритм, предложенный в работе [22]. Этот алгоритм обработки принимаемых сигналов включает в себя две основные операции.

Во-первых, параметры суммарных линейных (I/Q-дисбаланс) и нелинейных (в усилителе мощности) искажений сигнала, вызванных неидеальной работой блоков передатчика, оцениваются путем статистического анализа распределений комплексных амплитуд сигнала.

Во-вторых, «мягкие» метрики (LLR, *аббр. от англ.* Log-Likelihood Ratio), используемые в LDPC-декодере, вычисляются при демодуляции сигналов с учетом корректировки опорного модуляционного созвездия. Статистическая оценка искажений из-за I/Q-дисбаланса и нелинейности усилителя мощности проводилась на основе построения диаграмм рассеяния полученных комплексных амплитуд QAM-сигналов.

Пошаговое описание этапов обработки сигналов в радиоприемнике для компенсации искажений в радиопередатчике приведено ниже.

Шаг 1. Используя априорное знание структуры модуляционного созвездия (например, 16-QAM созвездия), в комплексной области принимаемого сигнала создается однородная двумерная сетка из ячеек достаточно малого размера. Путем тестирования предложенного алгоритма выявлено, что размер элементарных ячеек должен составлять не более 0,05 от максимальной амплитуды сигнала.

Шаг 2. Для каждой ячейки созданной двумерной сетки вычисляется количество символов принятого модулированного сигнала, попадающих в область этой ячейки.

Шаг 3. Производится обработка полученного двумерного массива счетчика символов (гистограмм) принятого модулированного сигнала, попадающих в области ячеек, путем двумерной фильтрации с гауссовским ядром. Данная обработка аналогична двумерной обработке изображений для борьбы с зашумленностью или после его сжатия. После двумерной фильтрации исходных данных производится поиск локальных максимумов, которые затем используются как новые точки опорного модуляционного созвездия. При практической реализации двумерной фильтрации радиус окна для гауссовского фильтра подбирался путем нескольких последовательных циклов обработки по критерию совпадения количества полученных максимумов с количеством точек в модуляционном созвездии принимаемого сигнала (например, до 16 точек для 16-QAM модуляции).

Шаг 4. Координаты точек, соответствующих локальным максимумам полученного двумерного массива, используются в качестве опорных в процессе демодуляции принятых сигналов при вычислении евклидовых расстояний в алгоритмах демодуляции сигналов для более точного вычисления «мягких» LLR-метрик (для каждого принятого бита), которые подаются на LDPC-декодер приемника.

Примеры гистограмм принимаемых сигналов с модуляцией 16-QAM, полученные до и после применения двумерной гауссовской фильтрации, показаны на рисунках 8a и 8b, соответственно.

Эффективность предложенного адаптивного алгоритма демодуляции и декодирования с компенсацией в радиоприемнике линейных и нелинейных искажений сигналов передатчиком была проверена путем обработки нескольких серий экспериментальных измерений характеристик созданного прототипа приемо-передающего оборудования. Наиболее эффективным оказалось применение данного алгоритма для демодуляции и декодирования сигналов с модуляцией 16-QAM и высокой скоростью кодирования 3/4.

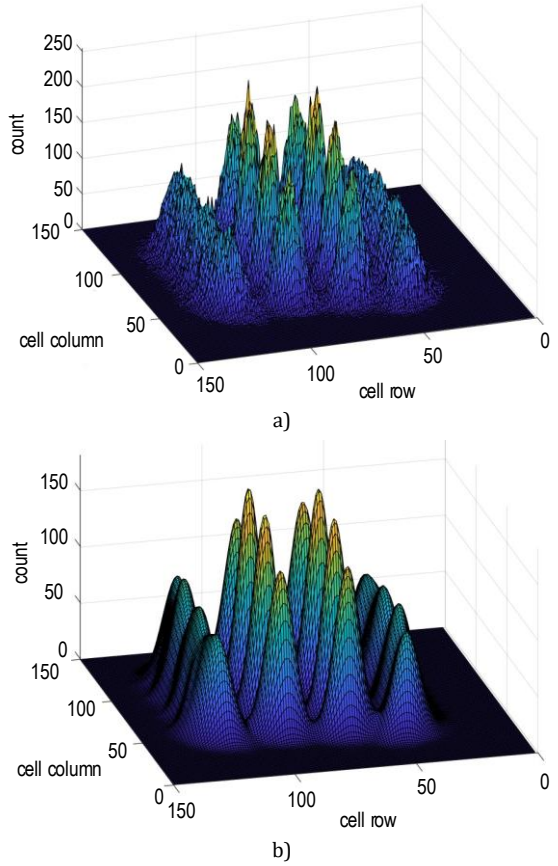


Рис. 8. Примеры гистограмм сигналов с модуляцией 16-QAM до (a) и после (b) применения двумерной гауссовской фильтрации

Fig. 8. Examples of Histograms of Signals with 16-QAM Modulation before (a) and after (b) the Two-Dimensional Gaussian Filtering Application

Результаты измерений вероятностей блоковых ошибок в 11 сериях экспериментов в случае использования адаптивного алгоритма компенсации и без него приведены в таблице 10, из которой видно, что применение предложенного адаптивного алгоритма при одинаковых условиях передачи позволило снизить вероятность пакетных ошибок приблизительно в два раза. По сравнению с известными схемами предварительного предискажения сигналов в радиопередатчике предлагаемый самообучающийся алгоритм демонстрирует гораздо меньшую сложность и поэтому может быть рекомендован для применения в приемо-передающем оборудовании небольших и дешевых релейных станций.

ТАБЛИЦА 10. Результаты измерений вероятностей блоковых ошибок

TABLE 10. Measurement Results of Block Error Probabilities

Номер эксперимента	BLER	
	без адаптации	с адаптацией
0	0,080	0,053
1	0,147	0,062
2	0,089	0,067
3	0,098	0,049
4	0,076	0,053
5	0,111	0,058
6	0,124	0,053
7	0,116	0,071
8	0,102	0,049
9	0,156	0,071
10	0,147	0,053
Ср. по всем экспериментам	0,113	0,058

Заключение

В ходе исследований характеристик разработанного прототипа приемо-передающего оборудования была продемонстрирована скорость передачи данных до 2 Гбит/с на расстояние 100 м и возможность уверенной передачи со скоростью до 500 Мбит/с на расстояние до 300 м с вероятностями пакетных ошибок менее 5 %. В то же время было выявлено, что используемые в РЧ-блоке прототипы микросхем HMC6000 и HMC6001 компании Hittite Microwave Corporation обладают высоким уровнем фазовых шумов, I/Q-дисбалансом около 8–10 % и при максимальных мощностях передатчика (10–12 дБм) достаточно большими нелинейными искажениями. Для борьбы с фазовыми шумами был разработан и реализован специальный алгоритм, компенсирующий линейный тренд фазовых набегов в радиоприемнике с помощью применения периодических коротких пилотных символов, используемых в стандарте IEEE 802.11ad. Для компенсации I/Q-дисбаланса модулятора/демодулятора и нелинейных искажений усилителя мощности передатчика был разработан и реализован адаптивный алгоритм демодуляции и декодирования (адаптивный LDPC-декодер), позволивший повысить эффективность передачи сигналов с модуляцией 16-QAM и уменьшить вероятность пакетных ошибок при пороговых уровнях SNR приблизительно в два раза, доведя их до приемлемого уровня.

Исследования характеристик прототипа приемо-передающего оборудования позволили также наметить план дальнейших работ по улучшению его характеристик с целью достижения более высоких скоростей передачи данных на большие расстояния. В частности, для улучшения характеристик прототипа планируется использовать новые РЧ-микросхемы с лучшими характеристиками, которые, согласно спецификациям, имеют меньший уровень фазовых шумов и меньшие нелинейные искажения. Кроме того, для передачи сигналов с модуляциями высокого порядка (64- и 256-QAM) возможно применение более сложных цифровых алгоритмов компенсации нелинейных искажений, в том числе и разрабатываемых в настоящее время специально для систем связи 5G NR в миллиметрового и субтерагерцового диапазонов длин волн.

Список источников

1. Rappaport T.S., Sun S., Mayzus R., Zhao H., Azar Y., Wang K., et al. Millimeter Wave Mobile Communications for 5G Cellular: It Will Work! // IEEE Access. 2013. Vol. 1. PP. 335–349. DOI:10.1109/ACCESS.2013.2260813

2. Boccardi F., Heath R.W., Lozano A., Marzetta T.L., Popovski P. Five disruptive technology directions for 5G // IEEE Communications Magazine. 2014. Vol. 52. Iss. 2. PP. 74–80. DOI:10.1109/MCOM.2014.6736746

3. Sakaguchi K., Haustein T., Barbarossa S., STRINATI E.C., Clemente A., DESTINO G., et al. Where, When, and How mmWave is Used in 5G and Beyond // IEICE Transactions on Electronics. 2017. Vol. E100-C. Iss. 10. PP. 790–808. DOI:10.1587/transele. E100.C.790

4. Liu D., Gaucher B., Pfeiffer U., Grzyb J. Advanced Millimeter-wave Technologies: Antennas, Packaging and Circuits. John Wiley & Sons, 2009. 832 p.

5. Perahia E., Cordeiro C., Park M., Yang L.L. IEEE 802.11ad: Defining the Next Generation Multi-Gbps Wi-Fi // Proceedings of the 7th IEEE Consumer Communications and Networking Conference (CCNC IEEE, Las Vegas, USA, 9–12 January 2010). IEEE, 2010. DOI:10.1109/CCNC.2010.5421713

6. Nitsche T., Cordeiro C., Flores A.B., Knightly E.W., Perahia E., Widmer J.C. IEEE 802.11ad: directional 60 GHz communication for multi-Gigabit-per-second Wi-Fi // IEEE Communications Magazine. 2014. Vol. 52. Iss. 12. PP. 132–141. DOI:10.1109/ MCOM.2014.6979964

7. 8802-11:2012/Amd.3:-2014 - ISO/IEC/IEEE. International Standard for Information technology--Telecommunications and information exchange between systems--Local and metropolitan area networks--Specific requirements-Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment 3: Enhancements for Very High Throughput in the 60 GHz Band (adoption of IEEE Std 802.11ad-2012). IEEE, 2014. DOI:10.1109/IEEESTD.2014.6774849

8. Ghasempour Y., da Silva C.R.C.M., Cordeiro C., Knightly E.W. IEEE 802.11ay: Next-Generation 60 GHz Communication for 100 Gb/s Wi-Fi // IEEE Communications Magazine. 2017. Vol. 55. Iss. 12. PP. 186–192. DOI:10.1109/MCOM.2017.1700393

9. Da Silva C.R.C.M., Lomayev A., Chen C., Cordeiro C. Analysis and Simulation of the IEEE 802.11ay Single-Carrier PHY // Proceedings of the International Conference on Communications (ICC, Kansas City, USA, 20–24 May 2018). IEEE, 2018. DOI:10.1109/ ICC.2018.8422532

10. 802.11ay-2021. IEEE Standard for Information Technology – Telecommunications and Information Exchange between Systems Local and Metropolitan Area Networks--Specific Requirements Part 11: Wireless LAN Medium Access Control (MAC)

and Physical Layer (PHY) Specifications Amendment 2: Enhanced Throughput for Operation in License-exempt Bands above 45 GHz. IEEE, 2021. DOI:10.1109/IEEESTD.2021.9502046

11. Dahlman E., Parkvall S., Skold J. 5G NR: The Next Generation Wireless Access Technology. Academic Press, 2018. DOI:10.1016/C2017-0-01347-2

12. Maltsev A., Lomayev A., Pudeyev A., Bolotin I., Bolkhovskaya O., Seleznev V. Millimeter-wave Toroidal Lens-Array Antennas Experimental Measurements // Proceedings of the International Symposium on Antennas and Propagation & USNC/URSI National Radio Science Meeting (Boston, USA, 08–13 July 2018). IEEE, 2018. PP. 607–608. DOI:10.1109/APUSNCURSINRSM. 2018.8608633

13. Bolkhovskaya O., Maltsev A., Seleznev V., Bolotin I. Cost-Efficient RAA Technology for Development of the High-Gain Steerable Antennas for mmWave Communications // In: Tallón-Ballesteros A.J., Chen C.H. (ed.) Machine Learning and Artificial Intelligence. Vol. 332. IOS Press, 2020. PP. 346–353. DOI:10.3233/FAIA200800

14. Yong S.-K., Xia P, Valdes-Garcia A. 60GHz Technology for Gbps WLAN and PAN: From Theory to Practice. John Wiley & Sons, 2011. 296 p.

15. Shabany M, Gulak P.G. Efficient Compensation of the Nonlinearity of Solid-State Power Amplifiers Using Adaptive Sequential Monte Carlo Methods // IEEE Transactions on Circuits and Systems I: Regular Papers. 2008. Vol. 55. Iss. 10. PP. 3270–3283. DOI:10.1109/TCSI.2008.925376

16. Bhat S., Chockalingam A. Compensation of power amplifier nonlinear distortion in spatial modulation systems // Proceedings of the 17th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC, Edinburgh, UK, 03–06 July 2016). IEEE, 2016. DOI:10.1109/SPAWC.2016.7536802

17. Maltsev A., Shikov A., Pudeev A., Kim S., Yang S. A Method for Power Amplifier Distortions Compensation at the RX Side for the 5G NR Communication Systems // In: Tallón-Ballesteros A.J. (ed.) Proceedings of CECNet 2022. Vol. 363. IOS Press, 2022. PP. 119–129. DOI:10.3233/FAIA220526

18. Wyglinski A.M., Getz R., Collins T., Pu D. Software-Defined Radio for Engineers. Artech House, 2018. 378 p.

19. Levanen T., Tervo O., Pajukoski K., Renfors M., Valkama M. Mobile Communications Beyond 52.6 GHz: Waveforms, Numerology, and Phase Noise Challenge // IEEE Wireless Communications. 2021. Vol. 28. Iss. 1. PP. 128–135. DOI:10.1109/MWC.001.2000185

20. Qi Y., Hunukumbure M., Nam H., Yoo H., Amuru S. On the Phase Tracking Reference Signal (PT-RS) Design for 5G New Radio (NR) // Proceedings of the 88th Vehicular Technology Conference (VTC-Fall, Chicago, USA, 27–30 August 2018). IEEE, 2018. DOI:10.1109/VTCFall.2018.8690852

21. Maltsev A., Pudeev A., Kim S., Yang S., Choi S., Myung S. Phase Tracking Sequences for 5G NR in 52.6–71 GHz Band: Design and Analysis // In: Tallón-Ballesteros A.J. (ed.) Proceedings of CECNet 2021. Vol. 345. IOS Press, 2021. PP. 268–282. DOI:10.3233/FAIA210412

22. Ermolaev G.A., Bolkhovskaya O.V., Maltsev A.A. Advanced Approach for TX Impairments Compensation Based on Signal Statistical Analysis at the RX Side // Proceedings of the Wave Electronics and its Application in Information and Telecommunication Systems (WECONF, St. Petersburg, Russia, 31 May 2021–04 June 2021). IEEE, 2021. DOI:10.1109/WECONF51603.2021.9470687

References

1. Rappaport T.S., Sun S., Mayzus R., Zhao H., Azar Y., Wang K., et al. Millimeter Wave Mobile Communications for 5G Cellular: It Will Work! *IEEE Access*. 2013;1:335–349. DOI:10.1109/ACCESS.2013.2260813

2. Boccardi F., Heath R.W., Lozano A., Marzetta T.L., Popovski P. Five disruptive technology directions for 5G. *IEEE Communications Magazine*. 2014;52(2):74–80. DOI:10.1109/MCOM.2014.6736746

3. Sakaguchi K., Haustein T., Barbarossa S., STRINATI E.C., Clemente A., DESTINO G., et al. Where, When, and How mmWave is Used in 5G and Beyond. *IEICE Transactions on Electronics*. 2017;E100-C(10):790–808. DOI:10.1587/transele.E100.C.790

4. Liu D., Gaucher B., Pfeiffer U., Grzyb J. *Advanced Millimeter-wave Technologies: Antennas, Packaging and Circuits*. John Wiley & Sons; 2009. 832 p.

5. Perahia E., Cordeiro C., Park M., Yang L.L. IEEE 802.11ad: Defining the Next Generation Multi-Gbps Wi-Fi. *Proceedings of the 7th IEEE Consumer Communications and Networking Conference, CCNC IEEE, 9–12 January 2010, Las Vegas, USA*. IEEE; 2010. DOI:10.1109/CCNC.2010.5421713

6. Nitsche T., Cordeiro C., Flores A.B., Knightly E.W., Perahia E., Widmer J.C. IEEE 802.11ad: directional 60 GHz communication for multi-Gigabit-per-second Wi-Fi. *IEEE Communications Magazine*. 2014;529120:132–141. DOI:10.1109/MCOM.2014.6979964

7. 8802-11:2012/Amd.3:-2014 - ISO/IEC/IEEE. *International Standard for Information technology--Telecommunications and information exchange between systems--Local and metropolitan area networks--Specific requirements-Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment 3: Enhancements for Very High Throughput in the 60 GHz Band (adoption of IEEE Std 802.11ad-2012)*. IEEE; 2014. DOI:10.1109/IEEESTD.2014.6774849

8. Ghasempour Y., da Silva C.R.C.M., Cordeiro C., Knightly E.W. IEEE 802.11ay: Next-Generation 60 GHz Communication for 100 Gb/s Wi-Fi. *IEEE Communications Magazine*. 2017;55(12):186–192. DOI:10.1109/MCOM.2017.1700393

9. Da Silva C.R.C.M., Lomayev A., Chen C., Cordeiro C. Analysis and Simulation of the IEEE 802.11ay Single-Carrier PHY. *Proceedings of the International Conference on Communications, ICC, 20–24 May 2018, Kansas City, USA*. IEEE; 2018. DOI:10.1109/ICC. 2018.8422532

10. 802.11ay-2021. *IEEE Standard for Information Technology – Telecommunications and Information Exchange between Systems Local and Metropolitan Area Networks--Specific Requirements Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment 2: Enhanced Throughput for Operation in License-exempt Bands above 45 GHz*. IEEE; 2021. DOI:10.1109/IEEESTD.2021.9502046

11. Dahlman E., Parkvall S., Skold J. *5G NR: The Next Generation Wireless Access Technology*. Academic Press; 2018. DOI:10.1016/C2017-0-01347-2

12. Maltsev A., Lomayev A., Pudeyev A., Bolotin I., Bolkhovskaya O., Seleznev V. Millimeter-wave Toroidal Lens-Array Antennas Experimental Measurements. *Proceedings of the International Symposium on Antennas and Propagation & USNC/URSI National Radio Science Meeting, 08–13 July 2018, Boston, USA*. IEEE; 2018. p.607–608. DOI:10.1109/APUSNCURSINRSM.2018.8608633

13. Bolkhovskaya O., Maltsev A., Seleznev V., Bolotin I. Cost-Efficient RAA Technology for Development of the High-Gain Steerable Antennas for mmWave Communications. *In: Tallón-Ballesteros A.J., Chen C.H. (ed.) Machine Learning and Artificial Intelligence. vol.332*. IOS Press; 2020. p.346–353. DOI:10.3233/FAIA200800

14. Yong S.-K., Xia P, Valdes-Garcia A. *60GHz Technology for Gbps WLAN and PAN: From Theory to Practice*. John Wiley & Sons; 2011. 296 p.

15. Shabany M, Gulak P.G. Efficient Compensation of the Nonlinearity of Solid-State Power Amplifiers Using Adaptive Sequential Monte Carlo Methods. *IEEE Transactions on Circuits and Systems I: Regular Papers*. 2008; 55(10):3270–3283. DOI:10.1109/TCSI.2008.925376

16. Bhat S., Chockalingam A. Compensation of power amplifier nonlinear distortion in spatial modulation systems. *Proceedings of the 17th International Workshop on Signal Processing Advances in Wireless Communications, SPAWC, 03–06 July 2016, Edinburgh, UK*. IEEE; 2016. DOI:10.1109/SPAWC.2016.7536802

17. Maltsev A., Shikov A., Pudeev A., Kim S., Yang S. A Method for Power Amplifier Distortions Compensation at the RX Side for the 5G NR Communication Systems. *In: Tallón-Ballesteros A.J. (ed.) Proceedings of CECNet 2022. vol.363*. IOS Press; 2022. p.119–129. DOI:10.3233/FAIA220526

18. Wyglinski A.M., Getz R., Collins T., Pu D. *Software-Defined Radio for Engineers*. Artech House; 2018. 378 p.

19. Levanen T., Tervo O., Pajukoski K., Renfors M., Valkama M. Mobile Communications Beyond 52.6 GHz: Waveforms, Numerology, and Phase Noise Challenge. *IEEE Wireless Communications*. 2021;28(1):128–135. DOI:10.1109/MWC.001.2000185.

20. Qi Y., Hunukumbure M., Nam H., Yoo H., Amuru S. On the Phase Tracking Reference Signal (PT-RS) Design for 5G New Radio (NR). *Proceedings of the 88th Vehicular Technology Conference, VTC-Fall, 27–30 August 2018, Chicago, USA*. IEEE; 2018. DOI:10.1109/VTCFall.2018.8690852

21. Maltsev A., Pudeev A., Kim S., Yang S., Choi S., Myung S. Phase Tracking Sequences for 5G NR in 52.6–71 GHz Band: Design and Analysis. *In: Tallón-Ballesteros A.J. (ed.) Proceedings of CECNet 2021. vol.345.* IOS Press; 2021. p.268–282. DOI:10.3233/FAIA210412

22. Ermolaev G.A., Bolkhovskaya O.V., Maltsev A.A. Advanced Approach for TX Impairments Compensation Based on Signal Statistical Analysis at the RX Side. *Proceedings of the Wave Electronics and its Application in Information and Telecommunication Systems, WECONF, 31 May 2021–04 June 2021, St. Petersburg, Russia*. IEEE; 2021. DOI:10.1109/WECONF51603.2021.9470687

Статья поступила в редакцию 29.12.2022; одобрена после рецензирования 31.03.2023; принята к публикации 04.04.2023.
The article was submitted 29.12.2022; approved after reviewing 31.03.2023; accepted for publication 04.04.2023.

Информация об авторах:

БОЛХОВСКАЯ Олеся Викторовна
кандидат физико-математических наук, доцент, доцент кафедры статистической радиофизики и мобильных систем связи Национального исследовательского Нижегородского государственного университета им. Н.И. Лобачевского
https://orcid.org/0000-0002-6679-9295

ЕРМОЛАЕВ Григорий Александрович
аспирант кафедры статистической радиофизики и мобильных систем связи Национального исследовательского Нижегородского государственного университета им. Н.И. Лобачевского
https://orcid.org/0000-0003-4213-953X

ТРУШКОВ Сергей Николаевич
аспирант кафедры статистической радиофизики и мобильных систем связи Национального исследовательского Нижегородского государственного университета им. Н.И. Лобачевского
https://orcid.org/0000-0002-5599-7157

МАЛЬЦЕВ Александр Александрович
доктор физико-математических наук, профессор, заведующий кафедрой статистической радиофизики и мобильных систем связи Национального исследовательского Нижегородского государственного университета им. Н.И. Лобачевского
https://orcid.org/0000-0001-8694-0033

Научная статья
УДК 519.872
DOI:10.31854/1813-324X-2023-9-2-40-46

Рекурсивный подбор параметров гиперэкспоненциальных распределений при аппроксимации распределений с «тяжелыми хвостами»

Марина Анатольевна Буранова, m.buranova@psuti.ru

Вячеслав Григорьевич Карташевский, v.kartashevskiy@psuti.ru

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

Аннотация: *Известно, что многие величины, определяющие сетевые характеристики функционирования инфокоммуникационной сети, имеют распределения вероятностей с «тяжелыми хвостами», которые могут оказать существенное влияние на производительность сети. Модели с распределениями, имеющими «тяжелый хвост», как правило, трудно исследовать. Анализ можно упростить с использованием аппроксимации распределения с «тяжелым хвостом» гиперэкспоненциальным распределением (конечной смесью экспонент). В работе приведен алгоритм расчета параметров компонент гиперэкспоненциального распределения, который основан на рекурсивном подборе параметров. Данный алгоритм позволяет анализировать различные модели очередей, включая G/G/1. Показано, что рассматриваемый подход наиболее целесообразно применять для аппроксимации монотонно убывающих распределений, имеющих «тяжелый хвост». Приведены примеры аппроксимации распределений Парето и Вейбулла.*

Ключевые слова: *гиперэкспоненциальное распределение, распределение с «тяжелым хвостом», рекурсивный подбор, системы массового обслуживания*

Ссылка для цитирования: Буранова М.А., Карташевский В.Г. Рекурсивный подбор параметров гиперэкспоненциальных распределений при аппроксимации распределений с «тяжелыми хвостами» // Труды учебных заведений связи. 2023. Т. 9. № 2. С. 40–46. DOI:10.31854/1813-324X-2023-9-2-40-46

Recursive Selection of Hyperexponential Distributions in Approximation of Distributions with "Heavy Tails"

Marina Buranova, m.buranova@psuti.ru

Vyacheslav Kartashevskiy, v.kartashevskiy@psuti.ru

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

Abstract: *It is known that many quantities that determine the network characteristics of the functioning of an infocommunication network have probability distributions with "heavy tails", which can have a significant impact on network performance. Models with heavy-tailed distributions tend to be difficult to analyze. The analysis can be simplified by using an algorithm to approximate a heavy-tailed distri-bution by a hyperexponential distribution (a*

finite mixture of exponentials). The paper presents a algorithm for calculating the parameters of the hyperexponential distribution components, which is based on a recursive selection of parameters. This algorithm allows you to analyze various models of queues, including G/G/1. It is shown that the approach under consideration is applicable to the approxi-mation of monotonically decreasing distributions, including those with a "heavy tail". Examples of approximation of Pareto and Weibull distributions are given.

Keywords: *hyperexponential distribution, distribution with a "heavy tail", recursive selection, queuing systems*

For citation: Buranova M., Kartashevskiy V. Recursive Selection of Hyperexponential Distributions in Approximation of Distributions with "Heavy Tails". *Proc. of Telecom. Universities.* 2023;9(2):40–46. (in Russ.) DOI:10.31854/1813-324X-2023-9-2-40-46

Введение

При анализе параметров функционирования современных инфокоммуникационных сетей одной из наиболее важных задач является разработка моделей, которые могут учитывать влияние особенностей обрабатываемого трафика. При этом традиционно анализ осуществляется с использованием методов теории массового обслуживания [1]. В качестве моделей систем обработки трафика очень часто используют систему $M/M/1$. В то же время известно, что современные потоки не обладают свойствами простейшего потока, для них характерно наличие фрактальных свойств, обусловленных, в частности, наличием «тяжелых хвостов» у распределений случайных значений интервалов времен между пакетами и интервалов обработки пакетов [2, 3]. Это требует разработки новых подходов к анализу систем обработки трафика, основанных на системах массового обслуживания с произвольными распределениями интервалов времени между пакетами и интервалов времени обработки пакетов, то есть систем $G/G/1$. Существуют разные модели системы $G/G/1$ и, пожалуй, одной из наиболее популярных является модель, основанная на использовании гиперэкспоненциального распределения. При этом систему $G/G/1$ аппроксимируют системой $H_l/H_k/1$, где символы H_l и H_k обозначают гиперэкспоненциальное распределение с числом экспонент l и k [4, 5]. Плотность вероятностей распределения, например для H_l, записывается в виде:

$$h(t) = \sum_{i=1}^{l} p_i \lambda_i e^{-\lambda_i t}, \qquad (1)$$

где p_i и λ_i – вес и параметр экспоненциальной компоненты; $\sum_{i=1}^{l} p_i = 1$.

При использовании такого подхода для модели $H_l/H_k/1$ задача сводится к определению числа экспонент смеси и параметров каждой экспоненциальной компоненты [5, 6]. Существуют различные методы определения данных параметров, в основном применительно к системе $H_2/H_2/1$, например, использование EM-алгоритма, как показано в работах [7–8], а также по первым двум или трем моментам исходного распределения [9].

Для выбора необходимого количества компонент системы $H_l/H_k/1$ и определения параметров модели можно воспользоваться подходом, изложенным в [5, 6], где приведены некоторые примеры определения параметров системы $M/G/1$.

Рассмотрим алгоритм определения параметров гиперэкспоненциального распределения при решении задачи аппроксимации распределений с «тяжелыми хвостами». Последние могут обладать бесконечным средним и бесконечной дисперсией, являются разновидностью распределений с «длинным хвостом»; при этом «хвост» может длиться достаточно долго при сохранении конечного значения первых двух моментов. В основе рассматриваемого подхода лежит возможность аппроксимировать распределения вероятностей с «длинным хвостом» простыми распределениями с «коротким хвостом», например, набором экспоненциальных распределений.

Учитывая, что процессы, протекающие в инфокоммуникационных сетях, как правило, определяются на конечном интервале, то для анализа их функционирования можно использовать модели системы массового обслуживания (СМО) при условии, что распределения случайных величин рассматриваются на конечном интервале $[t_1, t_2]$ [5].

Интервал $[t_1, t_2]$ разбивается на несколько подинтервалов, число которых соответствует числу экспонент в смеси. Параметры гиперэкспоненциальных распределений определяются на данных интервалах последовательно, начиная с интервала для максимальных значений случайной величины, где определяются первоначальные значения параметров гиперэкспоненты. Эта процедура повторяется рекуррентно для всех составляющих смеси экспонент на всех рассматриваемых интервалах.

Пусть $F(t)$ – интегральная функция распределения вероятностей; $F^{(c)}(t)$ – дополнительная интегральная функция распределения или функция распределения хвоста, при этом $F^{(c)}(t) = 1 - F(t)$. Обратим внимание, что распределение вероятностей имеет «длинный хвост», т. е. $F^{(c)}(t)$ убывает медленнее, чем экспоненциально: справедливо $F^{(c)}(t)$ at^{-b} при $t \to \infty$, для случая, когда a и b – положительные константы.

Известно, что одно из наиболее характерных распределений с «длинным хвостом» – распределение Вейбулла – имеет $F^{(c)}(t)$ в виде:

$$F^{(c)}(t) = e^{-\left(\frac{t}{z}\right)^{\alpha}}, \quad (2)$$

где α и z – параметры распределения Вейбулла.

Очевидно, что $F^{(c)}(t)$ для распределения Вейбулла в (2) имеет «тяжелый хвост», если $\alpha < 1$.

Для гиперэкспоненциального распределения H_k, состоящего из смеси k экспоненциальных распределений, дополнительную интегральную функцию можно записать в виде:

$$H^{(c)}(t) = \sum_{i=1}^{k} p_i e^{-\lambda_i t}, \quad (3)$$

где $p_i \geq 0$ для всех i и $\sum_{i=1}^{k} p_i = 1$.

В [5, 6] показано, что для случая, когда интегральная функция F имеет полностью монотонную плотность, существуют гиперэкспоненциальные интегральные функции $F^{(n)}, n \geq 1$, вида:

$$F^{(n)}(t) = \sum_{i=1}^{k_n} p\left(1 - e^{-\lambda t}\right), t \geq 0, \quad (4)$$

с $\lambda \leq \infty$ и $p_{n1} + \ldots + p_{nk_n} = 1$ такие, что $F^{(n)} \Rightarrow F$ при $n \to \infty$.

Основная идея заключается в выборе некоторого интегрального распределения $F^{(n)}$ с конечным числом экспонент, аппроксимирующего исходное F. Число экспонент, дающее необходимую точность аппроксимации, определяется экспериментальным путем.

Точность аппроксимации

Точность аппроксимации может определяться на основе анализа дополнительных интегральных функций или плотностей распределения вероятностей исходного распределения и его аппроксимации. В качестве численных показателей достигнутой точности подгонки можно использовать абсолютную погрешность представления интегральной и дополнительной интегральной функции распределения. Для обеих функций абсолютная ошибка представляется как:

$$AE(F,t) = \left|H^{(c)}(t) - F^{(c)}(t)\right| = |H(t) - F(t)|. \quad (5)$$

Относительная ошибка для функции распределения и дополнительной функции распределения, записывается в виде:

$$\Re(F,t) = \frac{|H^c(t) - F^c(t)|}{\min\{F(t), F^c(t)\}}. \quad (6)$$

Рекурсивная процедура подбора параметров гиперэкспоненциального распределения

Рассмотрим рекурсивную процедуру для подгонки интегрального гиперэкспоненциального распределения $H_k(t)$ к исходному интегральному распределению $F(t)$ в области положительных значений, аналогично показанному в [5, 6].

H_k имеет дополнительное интегральное распределение (3), и связанная с ней плотность распределения вероятностей имеет вид:

$$h(t) = \sum_{i=1}^{k} p_i \lambda_i e^{-\lambda_i t}, t \geq 0, \quad (7)$$

где $\sum_{i=1}^{k} p_i = 1, \lambda_i > 0, p_i > 0$ для всех i.

Пусть экспоненциальные параметры λ_i в (7) удовлетворяют условию: $\lambda_1 < \ldots < \lambda_k$. Тогда компоненты с более высокими индексами имеют «хвосты», которые затухают быстрее. Идея данного алгоритма состоит в том, чтобы рекурсивно подбирать компоненты H_k парами, то есть начиная с пары (λ_1, p_1), затем переходя к (λ_2, p_2) и так далее.

Рассмотрим предложенный в [5, 6] алгоритм определения параметров гиперэкспоненциального распределения, когда в качестве примера аппроксимируемого распределения используется распределение Вейбулла, с ${F_W}^{(c)}(t)$ в виде (2). В этом случае процедура определения параметров гиперэкспоненциального распределения включает четыре этапа.

Этап 1. Определяется число k экспоненциальных компонентов и k аргументов, по которым будут сопоставляться квантили: $0 < c_k < c_{k-1} < \ldots < c_1$, а также определяется параметр b, где c_k – квантиль распределения, определяемый как временной отрезок, на котором рассчитываются λ_i и p_i.

Этап 2. Определяется λ_1 и p_1 так, чтобы соответствовать функции ${F_W}^{c}(t)$ при аргументах c_1 и bc_1. При решении уравнений:

$$p_1 e^{-\lambda_1 c_1} = {F_W}^{(c)}(c_1) = \exp\left(-\left(\frac{c_1}{z}\right)^{\alpha}\right), \quad (8)$$

$$p_1 e^{-\lambda_1 b c_1} = {F_W}^{(c)}(bc_1) = \exp\left(-\left(\frac{bc_1}{z}\right)^{\alpha}\right) \quad (9)$$

для p_1 и λ_1 предполагается, что c_1 , b , ${F_W}^{(c)}(c_1)$ и ${F_W}^{(c)}(bc_1)$ известны.

Основываясь на (8) и (9), можно вычислить p_1 и λ_1 по выражениям:

$$\lambda_1 = \frac{1}{(b-1)c_1} \ln\left({F_W}^{(c)}(c_1) / {F_W}^{(c)}(bc_1)\right) = \frac{1}{(b-1)c_1} \ln\left(\frac{\exp\left(-\left(\frac{c_1}{z}\right)^{\alpha}\right)}{\exp\left(-\left(\frac{bc_1}{z}\right)^{\alpha}\right)}\right), \quad (10)$$

$$p_1 = F_W{}^{(c)}(c_1)e^{\lambda_1 c_1} = \\ = \exp\left(-\left(\frac{c_1}{z}\right)^{\alpha}\right)\cdot\exp(\lambda_1 c_1). \quad (11)$$

Этап 3. Определяются параметры λ_i и p_i для i-ой компоненты смеси при $2 \le i \le k$:

$$F_{Wi}^{(c)}(c_i) = F_{W(i-1)}^{(c)}(c_{i-1}) - \sum_{j=1}^{i-1} p_j e^{-\lambda_j c_i}, \quad (12)$$

$$F_{Wi}^{(c)}(bc_i) = F_{W(i-1)}^{(c)}(bc_{i-1}) - \sum_{j=1}^{i-1} p_j e^{-\lambda_j bc_i}. \quad (13)$$

При этом для распределения Вейбулла легко получить параметры i-ой компоненты смеси в виде:

$$\lambda_i = \frac{1}{(b-1)c_1}\ln\left(F_{Wi}^{(c)}(c_i)/F_{Wi}^{(c)}(bc_i)\right) = \\ = \frac{1}{(b-1)c_i}\ln\left(\frac{\exp\left(-\left(\frac{c_{i-1}}{z}\right)^{\alpha}\right) - \sum_{j=1}^{i-1} p_j e^{-\lambda_j c_i}}{\exp\left(-\left(\frac{bc_{i-1}}{z}\right)^{\alpha}\right) - \sum_{j=1}^{i-1} p_j e^{-\lambda_j bc_i}}\right), \quad (14)$$

$$p_i = \left(\exp\left(-\left(\frac{c_{i-1}}{z}\right)^{\alpha}\right) - \sum_{j=1}^{i-1} p_j e^{-\lambda_j c_i}\right)\cdot e^{\lambda_i c_i}. \quad (15)$$

Этап 4. Определяется последняя пара параметров (λ_k, p_k):

$$p_k = 1 - \sum_{j=1}^{k-1} p_j, \quad (16)$$

и, учитывая (12), λ_k рассчитывается в виде:

$$\lambda_k = \frac{1}{c_k}\ln\left(\frac{p_k}{\exp\left(-\left(\frac{c_{i-1}}{z}\right)^{\alpha}\right) - \sum_{j=1}^{i-1} p_j e^{-\lambda_j c_i}}\right). \quad (17)$$

Рекурсивная процедура подбора параметров гиперэкспонент для распределения Парето

Известно, что функция распределения Парето имеет вид:

$$F_P(x) = 1 - \left(\frac{k}{x}\right)^d, d > 0, k > 0, x > 0,$$

тогда дополнительная интегральная функция распределения Парето может быть записана как:

$$F_P{}^c(t) = \left(\frac{k}{t}\right)^d.$$

Алгоритм рекурсивного подбора параметров для распределения Парето будет аналогичен подходу, показанному для распределения Вейбулла, при этом для определения λ_1 и p_1 следует воспользоваться выражением (18). Параметры i-ой компоненты смеси для распределения Парето определяются по формулам (20, 21), и последняя пара параметров (λ_k, p_k) определяется по формулам (22, 23):

$$\lambda_1 = \frac{1}{(b-1)c_1}\ln\left(F_P{}^{(c)}(c_1)/F_P{}^{(c)}(bc_1)\right) = \\ = \frac{1}{(b-1)c_1}\ln\left(\frac{\left(\frac{k}{c_1}\right)^d}{\left(\frac{k}{bc_1}\right)^d}\right), \quad (18)$$

$$\lambda_i = \frac{1}{(b-1)c_1}\ln\left(F_{Pi}^{(c)}(c_i)/F_{Pi}^{(c)}(bc_i)\right) = \\ = \frac{1}{(b-1)c_i}\ln\left(\frac{\left(\frac{k}{c_{i-1}}\right)^d - \sum_{j=1}^{i-1} p_j e^{-\lambda_j c_i}}{\left(\frac{k}{bc_{i-1}}\right)^d - \sum_{j=1}^{i-1} p_j e^{-\lambda_j bc_i}}\right), \quad (20)$$

$$p_i = \left(\left(\frac{k}{c_{i-1}}\right)^a - \sum_{j=1}^{i-1} p_j e^{-\lambda_j c_i}\right)\cdot e^{\lambda_i c_i}. \quad (21)$$

$$p_k = 1 - \sum_{j=1}^{k-1} p_j, \quad (22)$$

$$\lambda_k = \frac{1}{c_k}\ln\left(\frac{p_k}{\left(\frac{k}{c_{i-1}}\right)^d - \sum_{j=1}^{i-1} p_j e^{-\lambda_j c_i}}\right). \quad (23)$$

Примеры аппроксимации для распределений с «тяжелыми хвостами»

В качестве примеров рассмотрим распределение Вейбулла с двумя наборами параметров:

1) α = 0,8; z = 0,8865; средним значением m = 1, коэффициентом вариации V= 1,26; σ² = 1,6;

2) α = 0,6; z = 0,6646; средним значением m = 1; коэффициентом вариации V= 1,7; σ² = 3,09.

А также – распределение Парето с параметрами:

3) d = 2,2; k = 0,55; средним значением m = 1; коэффициентом вариации V= 1,5; σ² = 2,31.

Используемые значения коэффициента вариации в данных примерах показывают, что рассматриваемые распределения обладают «тяжелыми хвостами». Результаты аппроксимации при условии использования различного числа экспонент в смеси представлены на рисунке 1.

Для распределения Вейбулла результаты аппроксимации показывают, что данный алгоритм позволяет добиться необходимой точности за счет увеличения числа экспонент. При аппроксимации H_{20} функция распределения хвоста лежит значительно выше функции исходного распределения, что излишне его «утяжеляет». Достаточным для рассматриваемого случая можно принять результат аппроксимации H_6, поскольку он дает удовлетворительный уровень точности при небольшом числе экспонент.

На рисунке 2 представлены зависимости ошибки от времени для подгонки H_2, H_6, H_{10} и H_{20} согласно выражениям (5) и (6).

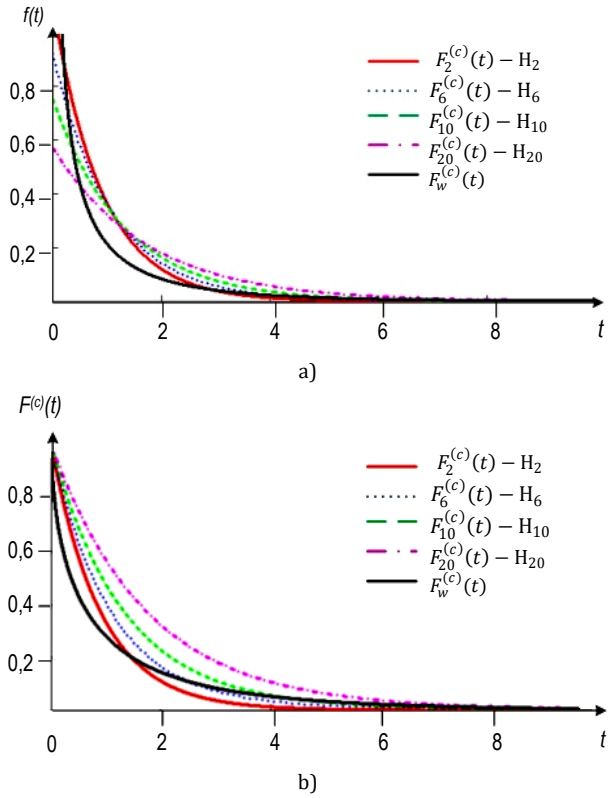


Рис. 1. Графики аппроксимации распределения Вейбулла: a) плотность; b) дополнительная интегральная функция

Fig. 1. Graphs of Weibull Distribution Approximation: a) Density; b) Complementary Cumulative Distribution Function

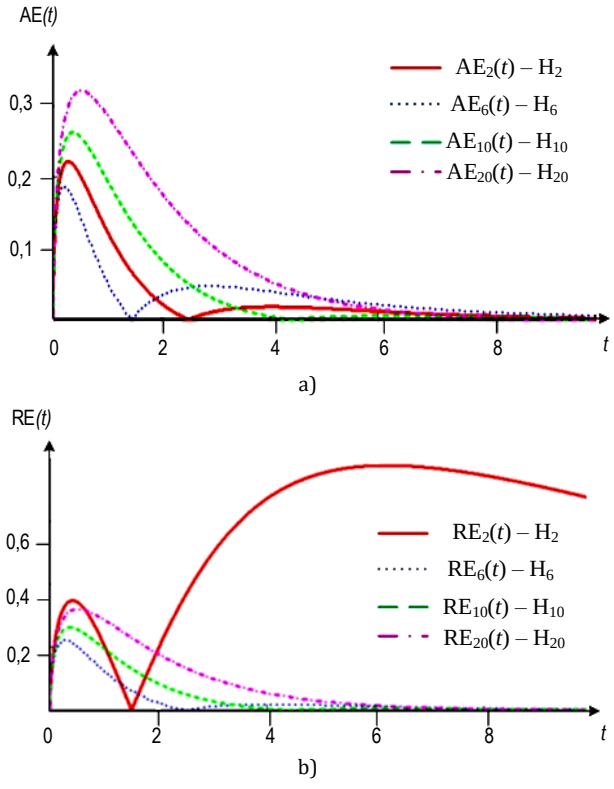


Рис. 2. Ошибка аппроксимации распределения Вейбулла: a) абсолютная; b) относительная

Fig. 2. Approximation Error of the Weibull Distribution: a) Absolute; b) Relative

Анализ результатов, представленных на рисунках 1 и 2, показывает, что наиболее точную аппроксимацию показывает гиперэкспоненциальное распределение с шестью экспонентами. Абсолютная ошибка аппроксимации составляет от 1,5 до 12 %; ошибка в 12 % соответствует согласно (5) точке максимального расхождения кривых.

Пример аппроксимации распределения Парето показан на рисунке 3. Анализ аппроксимации распределения Парето дает результаты, аналогичные полученным для распределения Вейбулла. Визуально наиболее близкими являются аппроксимации с двумя и шестью экспонентами.

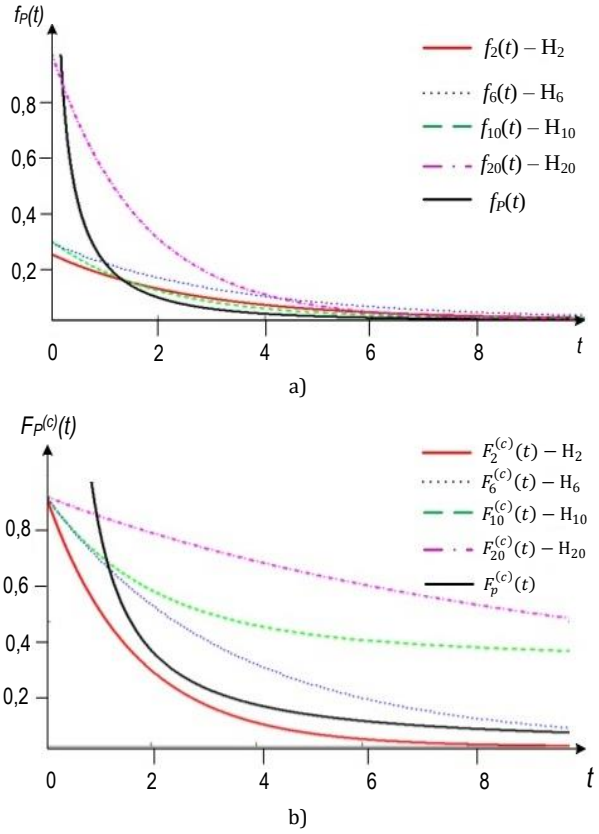


Рис. 3. Графики аппроксимации распределения Парето: a) плотность; b) дополнительная интегральная функция

Fig. 3. Graphs of Pareto Distribution Approximation: a) Density; b) Complementary Cumulative Distribution Function

Абсолютная и относительная ошибка для распределения Парето при аппроксимации двумя и шестью экспонентами показана на рисунке 4. При этом ошибка аппроксимации распределения Парето гиперэкспоненциальным распределением с шестью экспонентами составляет от 1,5 до 30%. 30% соответствует точке максимального расхождения кривых. В среднем ошибка не превышает 5% и для случая аппроксимации распределения Вейбулла и для случая аппроксимации распределения Парето.

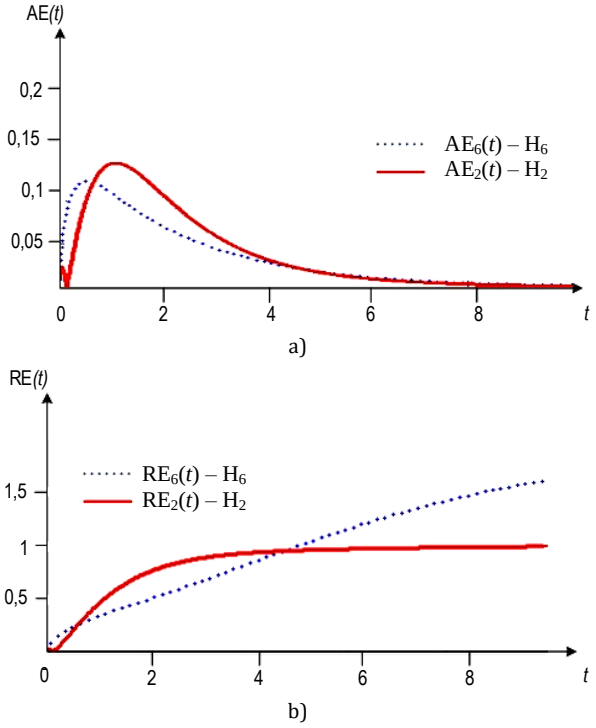


Рис. 4. Ошибка аппроксимации распределения Парето: а) абсолютная; b) относительная

Fig. 4. Approximation Error of the Pareto Distribution: a) Absolute; b) Relative

Особенностью данного алгоритма является то, что он не требует знания моментов распределения при его реализации. Поэтому его можно использовать, даже если моменты не существуют или неизвестны. Однако иногда бывает полезно вычислить несколько первых моментов исходного и аппроксимирующего распределений, чтобы оценить качество подгонки. В нашем случае, как было показано выше, качество подгонки было определено через оценивание точности с использованием выражений (5) и (6).

Заключение

В работе рассмотрен алгоритм определения параметров гиперэкспоненциальных распределений, применяемый для аппроксимации монотонно убывающих распределений, с использованием подхода, основанного на рекурсивном подборе параметров гиперэкспонент. Показано, что данный алгоритм может быть успешно использован для аппроксимации распределений из класса распределений с «тяжелыми хвостами».

Приведены примеры, показывающие, что алгоритм эффективен для аппроксимации распределений Парето и Вейбулла; для первого абсолютная ошибка аппроксимации составляет от 1,5 до 12 %, а для второго – от 0,2 до 30%. Такие значения ошибки не превышают значений, получаемых при использовании других методов аппроксимации, например, методов, основанных на определении первых 2-х или 3-х моментов исходного распределения. Полностью монотонные плотности распределений могут быть аппроксимированы гиперэкспоненциальными плотностями распределений с необходимой точностью.

Предложенный алгоритм позволяет с высокой точностью провести статистическую аппроксимацию любого распределения, в том числе распределений с тяжелыми хвостами, весовой суммой экспонент. Представленный подход позволяет решать задачи определения основных параметров функционирования систем G/G/1, например, таких как задержка и вариация задержки.

Установлено, что рассмотренный алгоритм определения параметров гиперэкспоненциальных распределений, с использованием подхода, основанного на рекурсивном подборе параметров гиперэкспонент, позволяет получить аппроксимацию исходного распределения с высокой точностью (ошибка не более 5 %).

Представляет интерес развитие данного подхода на модели систем, обрабатывающие трафик, обладающий фрактальными свойствами, что позволяет учесть корреляционные свойства трафика.

В дальнейших исследованиях предполагается дать сравнительный анализ точности и вычислительной сложности предложенного алгоритма с другими возможными алгоритмами аппроксимации.

Список источников

1. Клейнрок Л. Теория массового обслуживания. Пер. с англ. М.: Машиностроение, 1979. 432 с.

2. Шелухин О.И., Смольский С.М., Осин А.В. Самоподобие и фракталы. Телекоммуникационные приложения. М.: Физматлит, 2008. 368 с.

3. Kotz S., Johnson N.L., Read C.B. Encyclopedia of Statistical Sciences. Vol. 8. New York: Wiley, 1988. PP. 352–357.

4. Keilson J., Machihara F. Hyperexponential waiting time structure in hyperexponential H_K/H_L/1 system // Journal of the Operation Research Society of Japan. 1985. Vol. 28. Iss. 3. PP. 242–250. DOI:10.15807/jorsj.28.242

5. Feldmann A., Whitt W. Fitting mixtures of exponentials to long-tail distributions to analyze network performance models // Performance Evaluation. 1998. Vol. 31. Iss. 3–4. PP. 245–279. DOI:10.1016/S0166-5316(97)00003-5

6. Буранова М.А., Карташевский В.Г. Определение параметров гиперэкспоненциального распределения методом рекурсивного подбора // XXVII Международная научно-техническая конференция, посвященная 60-летию полетов в космос Ю.А. Гагарина и Г.С. Титова «Радиолокация, навигация, связь» (Воронеж, Россия, 28–30 сентября 2021). Воронеж: Издательский дом ВГУ, 2021. С. 43–54.

7. Королев В.Ю. ЕМ-алгоритм его модификации и их применение к задаче разделения смесей вероятностных распределений. Теоретический обзор. М.: ИПИ РАН, 2007. 94 с.

8. Buranova M., Ergasheva D., Kartashevskiy V. Using the EM-algorithm to Approximate the Distribution of a Mixture by Hyperexponents // Proceedings of the International Conference on Engineering and Telecommunication (EnT, Dolgoprudny, Russia, 20–21 November 2019). IEEE, 2019. DOI:10.1109/EnT47717.2019.9030551

9. Тарасов В.Н., Карташевский И.В. Определение среднего времени ожидания требований в управляемой системе массового обслуживания H2/H2/1 // Системы управления и информационные технологии. 2014. №3(57). С. 92–96.

References

1. Kleinrock L. *Queueing Systems. Vol. 1: Theory*. New York: Wiley-Interscience, 1975. 432 p.

2. Sheluhin O.I., Smolsky S.M., Osin A.V. *Self-Similarity and Fractals. Telecommunication Applications*. Moscow: Fizmatlit Publ.; 2008. 368 p. (in Russ.)

3. Kotz S., Johnson N.L., Read C.B. *Encyclopedia of Statistical Sciences.* New York: Wiley; 1988. vol.8. p.352–357.

4. Keilson J., Machihara F. Hyperexponential waiting time structure in hyperexponential H_K/H_L/1 system. *Journal of the Operation Research Society of Japan*. 1985;28(3):242–250. DOI:10.15807/jorsj.28.242

5. Feldmann A., Whitt W. Fitting mixtures of exponentials to long-tail distributions to analyze network performance models. *Performance Evaluation*. 1998;31(3–4):245–279. DOI:10.1016/S0166-5316(97)00003-5

6. Buranova M.A., Kartashevskii V.G. Determination of the Parameters of Hyperexponential Distribution by the Method of Recursive Selection. *Proceedings of the XXVIIth International Technical Conference on Radar, Navigation, Communications, 28–30 September 2021, Voronezh, Russia*. Voronezh: Voronezh State University Publ.; 2021. p.43–54. (in Russ.)

7. Korolyov V.Yu. *The EM Algorithm, Its Modifications, and Their Application to the Problem of Separating Mixtures of Probability Distributions. Theoretical Review*. Moscow: IPI RAN Publ.; 2007. 94 p. (in Russ.)

8. Buranova M., Ergasheva D., Kartashevskiy V. Using the EM-algorithm to Approximate the Distribution of a Mixture by Hyperexponents. *Proceedings of the International Conference on Engineering and Telecommunication, EnT, 20–21 November 2019, Dolgoprudny, Russia)*. IEEE; 2019. DOI:10.1109/EnT47717.2019.9030551

1. Tarasov V.N., Kartashevskij I.V. Determination of the average Waiting Time for Requirements in a Managed Mass Service System. *Sistemy upravleniya i informaci-onnye tekhnologii*. 2014;3(57):92–96. (in Russ.)

Статья поступила в редакцию 14.03.2023; одобрена после рецензирования 22.03.2023; принята к публикации 20.04.2023.
The article was submitted 14.03.2023; approved after reviewing 22.03.2023; accepted for publication 20.04.2023.

Информация об авторах:

БУРАНОВА Марина Анатольевна
доктор технических наук, доцент, доцент кафедры информационной безопасности, начальник управления организации учебного процесса Поволжского государственного университета телекоммуникаций и информатики
https://orcid.org/0000-0003-2986-8252

КАРТАШЕВСКИЙ Вячеслав Григорьевич
доктор технических наук, профессор, заведующий кафедрой информационной безопасности Поволжского государственного университета телекоммуникаций и информатики
https://orcid.org/0000-0003-1114-3966

Научная статья
УДК 621.39, 530.182
DOI:10.31854/1813-324X-2023-9-2-47-56

Квазисолитонный режим в многопролетной волоконно-оптической системе связи с применением оптических усилителей

Сергей Федорович Глаголев, Glagolev.Sergey@sut.ru
Сергей Эдуардович Доценко, Dotsenko.Sergey@sut.ru

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

Аннотация: *В работе рассмотрен метод поддержания квазисолитонного режима в многопролетной волоконно-оптической системе связи с использованием дискретных эрбиевых оптических усилителей, а также рамановских усилителей с распределенным усилением со встречной и двунаправленной накачкой. Для моделирования квазисолитонных волоконно-оптических систем связи использовалались программы OptiSystem 19. Результаты моделирования сопоставлены с теоретическими данными, продемонстрированы преимущества волоконно-оптических систем связи с усилителями Рамана и двунаправленной накачкой.*

Ключевые слова: *солитон, квазисолитонный режим, волоконно-оптическая система связи, эрбиевый оптический усилитель, оптический усилитель Рамана, встречная и двунаправленная накачка, OptiSystem*

Ссылка для цитирования: Глаголев С.Ф. Доценко С.Э. Квазисолитонный режим в многопролетной волоконно-оптической системе связи с применением оптических усилителей // Труды учебных заведений связи. 2023. Т. 9. № 2. С. 47–56. DOI:10.31854/1813-324X-2023-9-2-47-56

Quasi-Soliton Mode in a Multi-Span Fiber-Optic Communication System Using Optical Amplifiers

Sergey Glagolev, Glagolev.Sergey@sut.ru
Sergey Dotsenko, Dotsenko.Sergey@sut.ru

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

Abstract: *The paper considers a method for maintaining a quasi-soliton mode in a multi-span fiber-optic communication system using discrete Erbium optical amplifiers, as well as Raman amplifiers with distributed amplification with counter and bidirectional pumping. The OptiSystem 19 and OptiPerformer 19 programs were used to model quasi-soliton fiber-optic communication systems. The simulation results are compared with theoretical data, and the advantages of a fiber-optic communication system with a Raman amplifier and bidirectional pumping are demonstrated.*

Keywords: *soliton, quasi-soliton mode, Raman optical amplifier, OptiSystem 19, OptiPerformer 19*

For citation: Glagolev S., Dotsenko S. Quasi-Soliton Mode in a Multi-Span Fiber-Optic Communication System Using Optical Amplifiers. *Proc. of Telecom. Universities.* 2023;9(2):47–56. (in Russ.) DOI:10.31854/1813-324X-2023-9-2-47-56

Введение

Данная статья продолжает работы авторов [1, 2], которые посвящены применению дискретных оптических усилителей (ОУ) на основе эрбиевого оптического волокна (EDFA, *аббр. от англ.* Erbium Doped Fiber Amplifier) и распределенных усилителей Рамана (RA, *аббр. от англ*. Raman Amplifier) для поддержания квазисолитонного режима распространения сигналов в многопролетных волоконно-оптических системах связи (ВОСС). Кратко резюмируем полученные ранее результаты.

В одномодовом оптическом волокне (ОМОВ) без потерь могут распространяться, не меняя пиковой мощности P_m (ПМ) и полуширины T_0, оптические импульсы, имеющие форму гиперболического секанса $P(t) = P_m \operatorname{sech}(t/T_0)$. Такие импульсы называют фундаментальными солитонами или импульсами секансной формы. В реальном ОМОВ с потерями фундаментальные солитоны существовать не могут, так как ПМ оптических импульсов с увеличением расстояния уменьшается и ее становится недостаточно для поддержания солитонного режима, импульсы затухают и расширяются. Разбив волоконно-оптический линейный тракт (ВОЛТ) на несколько участков (пролетов), можно, используя оптическое усиление, добиться в каждом пролете многопролетной ВОСС поддержания с некоторой погрешностью средней ПМ и длительности секансных импульсов (квазисолитонов). Пролеты в таком ВОЛТ называют прозрачными, т. к. ПМ на входе и выходе пролетов одинаковы. Этот метод поддержания квазисолитонного режима в литературе называют «управление затуханием» [1–5].

В статье приводятся результаты исследований, необходимых для оптимального выбора длины $L_{пр}$ и количества пролетов N в многопролетной квазисолитонной ВОСС определенной длины $L = NL_{пр}$.

Напомним, что фундаментальный солитон может существовать только в ОМОВ без потерь и с аномальной дисперсией при полной компенсации хроматической дисперсии (ХД) в результате действия нелинейной фазовой самомодуляции (ФСМ). ХД характеризуется величиной β_2 дисперсии групповых скоростей (ДГС), а ФСМ – коэффициентом нелинейности γ. Это возможно только при определенном соотношении ПМ фундаментального солитона $P_{Фс}$ и его полуширины T_0, которая однозначно связана с длительностью секансного импульса t_u на уровне половины амплитуды [1–5]:

$$P_{Фс} = |\beta_2|/(\gamma \cdot T_0^2) \approx 3{,}11 \cdot |\beta_2|/(\gamma \cdot t_u^2). \qquad (1)$$

Многопролетная квазисолитонная ВОСС с дискретными ОУ EDFA

Схема многопролетной ВОСС с дискретными ОУ (рисунок 1) содержит два оконечных пункта с транспондерами ТР1 и ТР2, $N + 1$ пролетов длиной $L_{пр}$ и N линейных ОУ, которые расположены в усилительных пунктах (УП).

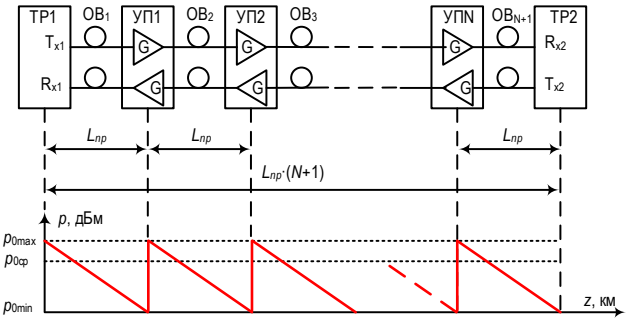


Рис. 1. Схема линейного тракта ВОСС с дискретными ОУ

Fig. 1. Diagram of the Linear Path of a Fiber-Optic Communication System with Discrete Optical Amplifiers

Приведем параметры, использованные для расчета и моделирования квазисолитонной ВОСС: скорость передачи – B = 10 Гбит/с; длина волны – λ = 1550 нм; длительность секансного импульса на уровне половины амплитуды – t_u = 0,2 бит (20 пс). Параметры ОВ DSF [6]: коэффициент затухания – α = 0,2 дБ/км (0,046 Нп/км) на длине волны λ = 1550 нм; коэффициент ХД – D_x = 1 пс/(км·нм); ДГС – $\beta_2 = -1{,}275$ пс/нм^2; крутизна дисперсионной характеристики – S_x = 0,085 пс/нм^2/км; эффективная площадь модового поля – A_{ef} = 41 мкм^2; нелинейный показатель преломления – $n_2 = 26 \cdot 10^{-21}$ м^2/Вт; коэффициент нелинейности – γ = 2,57 1/(Вт·км).

Проведем расчеты в соответствии с [1, 2]: канонической полуширины секансного импульса – $T_0 = t_u/1{,}763 = 11{,}34$ пс, мощности, необходимой для поддержания солитонного режима в ОВ без потерь по уравнению (1) $P_{Фс} = -\beta_2/(\gamma \cdot T_0^2) = 3{,}9$ мВт и дисперсионной длины $L_D = T_0^2/(-\beta_2) = 100$ км.

Схема имитационного моделирования многопролетной квазисолитонной ВОСС, созданная в программе OptiSystem 19 [7] (рисунок 2), включает оптический передатчик с амплитудной модуляцией и импульсами секансной формы для передачи логических «1», ВОЛТ с любым количеством одинаковых пролетов, оптический полосовой фильтр и фотоприемное устройство (*Rx*). Один прозрачный пролет содержит ОВ длиной $L_{пр}$ с затуханием $a_{пр} = \alpha L_{пр}$ и линейный оптический усилитель EDFA с усилением $G = a_{пр}$.

Для контроля оптических и электрических сигналов в схеме (см. рисунок 2) используются оптические измерители мощности (Optical Power Meter), осциллографы (Optical Time Domain Visualizer) и спектроанализаторы (Optical Spectrum Analyzer), электрические осциллографы (Oscilloscope Visualizer) и анализатор ошибок (BER Analyzer).

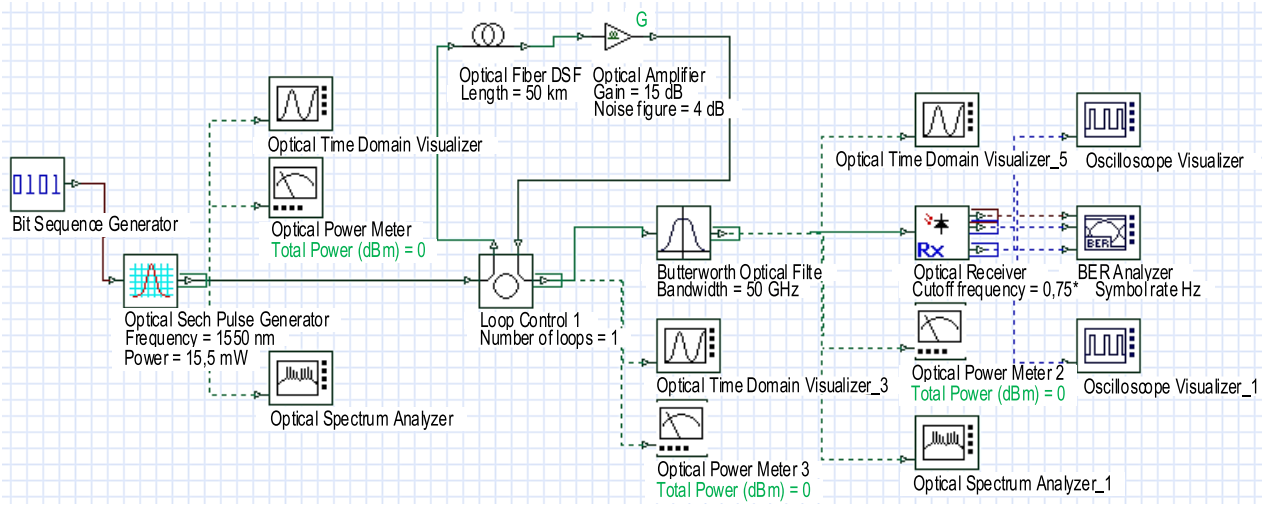


Рис. 2. Схема моделирования многопролетной квазисолитонной ВОСС с дискретными ОУ

Fig. 2. Simulation Scheme of a Multi-Span Quasi-Soliton Fiber-Optic Communication System with Discrete Optical Amplifiers

ТАБЛИЦА 1. Результаты расчетов и имитационного моделирования для ВОСС с дискретными линейными ОУ

TABLE 1. Results of Calculations and Measurements for a Fiber-Optic Communication System with Discrete Linear Optical Amplifiers

Установленные и расчетные величины							Результаты имитационного моделирования		
Количество пролетов	Длина ОМОВ, км	Эффективная длина ОМОВ, км	Коэффициент усиления ОУ, дБ	Расчетная входная ПМ, мВт	Длительный входной импульса, пс	Установка входной ПМ, мВт	Выходная ПМ, мВт	Длительность выходного импульса, пс	Q фактор
1	25	14,86	5	6,56	20	7,5	7,6	20	315
2	50	14,86	5	6,56	20	7,5	7,6	20	252
4	100	14,86	5	6,56	20	7,5	7,7	20	200
8	200	14,86	5	6,56	20	7,5	7,7	20	144
12	300	14,86	5	6,56	20	7,5	7,8	20	127
20	500	14,86	5	6,56	20	7,5	7,8	19,5	95
1	50	19,56	10	9,97	20	11	11,4	19,5	247
2	100	19,56	10	9,97	20	11	11,4	20	209
4	200	19,56	10	9,97	20	11	11,2	20	144
8	400	19,56	10	9,97	20	11	10,8	20	106
10	500	19,56	10	9,97	20	11	10,6	20,5	95
1	75	21,0	15	13,9	20	15,5	16,2	20	195
2	150	21,0	15	13,9	20	15,5	16	20	147
4	300	21,0	15	13,9	20	15,5	15,3–16	18–21	102
6	450	21,0	15	13,9	20	15,5	14-14,4	22–23	78
1	100	21,5	20	18,1	20	22	23	20	140
2	200	21,5	20	18,1	20	22	23	20	94
3	300	21,5	20	18,1	20	22	21,8–23,0	19,8–20,8	75
4	400	21,5	20	18,1	20	22	18,4–20,6	22–26	64

Исследования проводились для четырех длин пролетов 0,25; 0,5; 0,75 и 1,0 L_D, (25, 50, 75 и 100 км). Для каждой $L_{\text{пр}}$ рассчитывалась эффективная длина ОМОВ $L_{ef} = [1 - \exp(-\alpha \cdot L)]/\alpha$ (α в Нп/км), ПМ на входе пролета $P_{0m} = P_{\text{Фс}} \cdot L_{\text{пр}}/L_{ef}$ и коэффициент усиления $G = \alpha\, L_{np}$. Результаты расчетов и измерений приведены в таблице 1, а формы импульсов на выходе ВОСС на рисунке 3.

Из таблицы 1 и рисунка 3 следует, что при длинах пролета 0,25 и 0,5 L_D и общей длине до 500 км квазисолитонный режим сохраняется, джиттер отсутствует. При этом длительность секансного импульса и его амплитуда при изменении количества пролетов сохраняются в пределах 2,5 и 4 %, соответственно.

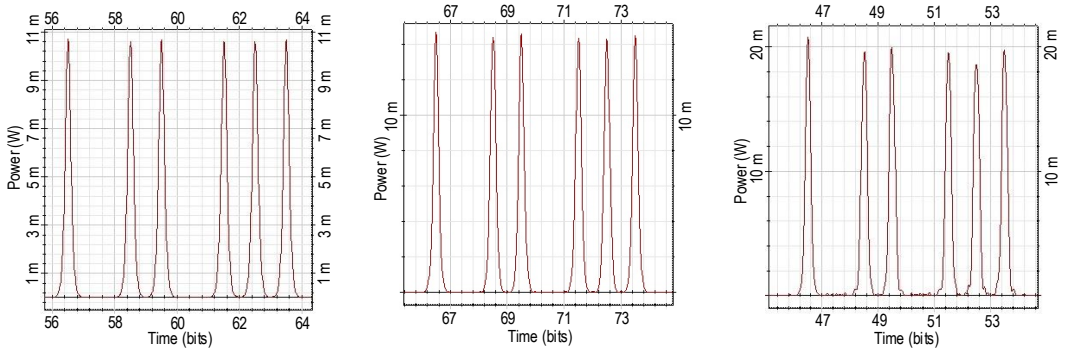


Рис. 3. Форма импульсов на выходах многопролетных квазисолитонных ВОСС: длина 50 км × 10 = 500 км (a); 75 км × 6 = 450 км (b); 100 км × 4 = 400 км (c)

Fig. 3. The Shape of Pulses at the Outputs of Multi-Span Quasi-Soliton Fiber-Optic Communication Systems: Length 50 km × 10 = 500 km (a); 75 km × 6 = 450 km (b); 100 km × 4 = 400 km (c)

При длине пролета 0,75 и 1,0 L_D и общей длине до 200 км квазисолитонный режим сохраняется, джиттер отсутствует. После 300 км наблюдается значительный джиттер, который приводит к колебаниям ПМ и длительности импульсов от импульса к импульсу в пределах 5 %.

Расчетные значения входных ПМ для поддержания квазисолитонного режима оказались недостаточными и реальные мощности (см. таблицу 1) были больше в 1,1–1,2 раза.

Многопролетная ВОСС с RA и встречной накачкой

Рассмотрим схему (рисунок 4) многопролетной ВОСС с RA, которые также называют ОУ вынужденного комбинационного рассеяния (ОУ ВКР) со встречной накачкой [8–10]. В оконечных пунктах ВОСС находятся транспондеры ТР1, ТР2. Для поддержания квазисолитонного режима в многопролетной ВОСС, с выхода каждого пролета в ОМОВ через направленный ответвитесь (НО) с технологией мультиплексирования в волновой области (WDM) подается навстречу сигналу непрерывное излучение от источника накачки (ИН).

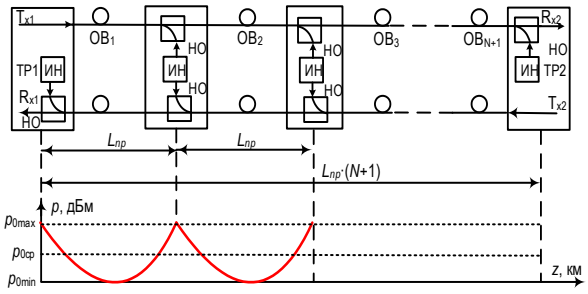


Рис. 4. Схема линейного тракта ВОСС с RA и встречной накачкой

Fig. 4. Diagram of the Linear Path of a Fiber-Optic Communication System with Counter-Pumping of the Raman

Рассмотрим процессы распространения секансных импульсов в одном пролете (на одном усилительном участке УУ) квазисолитонной ВОСС. Запишем дифференциальное уравнение для изменений ПМ $P_m(z)$ на малом участке dz в ОВ с усилительной способностью $g(z)$ и коэффициентом затухания α [1–5]:

$$\frac{dP_m(z)}{dz} = [g(z) - \alpha] \cdot P_m(z). \qquad (2)$$

На УУ длиной $L_{\text{пр}}$ усилительная способность $g(z)$ зависит от z и определяется уровнем накачки в этой точке. В конце УУ при $z = L_{\text{пр}}$ усилительная способность $g(z) = g_0$ и уровень мощности накачки максимальны. При $g(z) < \alpha$, ПМ на участке dz уменьшается, а при $g(z) > \alpha$ – возрастает.

Из рисунка 4 видно, что в начале пролета уровень сигнала с увеличением расстояния убывает, в связи с тем, что мощность накачки мала и преобладает затухание $g(z) < \alpha$. С увеличением расстояния z мощность накачки постепенно возрастает и сигнал, пройдя минимум, начинает возрастать, т. к. $g(z) > \alpha$. В конце пролета сигнал приобретает первоначальный уровень. Средний пиковый уровень сигнала в пролете должен быть равен пиковому уровню фундаментального солитона $P_{\text{Фс}}$.

В каждой точке ОВ невозможно компенсировать потери, но можно скомпенсировать общее затухание на УУ:

$$\int_0^{L_A} g(z) \cdot dz = \alpha \cdot L_{\text{пр}}. \qquad (3)$$

Пренебрегая истощением накачки, запишем упрощенное выражение для усилительной способности на УУ:

$$g(z) = g_0 \cdot \exp\left[-\alpha_p \cdot \left(L_{\text{пр}} - z\right)\right], \qquad (4)$$

где α_p – коэффициент затухания ОВ для накачки.

Решив (2) с учетом (3) и (4), запишем выражение (5) для расчета ПМ сигнала, которая обеспечивает необходимую усилительную способность g_0 и

при которой ПМ на выходе пролета будет соответствовать входной $P_{0m} = P_m(L_{пр})$ [1–5]:

$$P_m(z) = P_{0m} \cdot \exp\left\{\alpha \cdot \left[L_{пр} \cdot \left[\frac{\exp(\alpha_p \cdot z) - 1}{\exp(\alpha_p \cdot L_{пр}) - 1}\right] - z\right]\right\} = P_{0m} \cdot p(z), \quad (5)$$

где $p(z) = P_m(z)/P_{0m}$ – относительная ПМ в пролете.

В случае использования встречной накачки, ПМ входных импульсов P_{0m} должна быть больше мощности фундаментального солитона $P_{Фс}$ в ОМОВ без потерь [1–5]:

$$P_{0m} = P_{Фс}/p_0\,, p_0 = (1/L_{пр}) \cdot \int_0^{L_A} p(z) \cdot dz, \quad (6)$$

где p_0 – среднее значение относительной ПМ излучения в пролете.

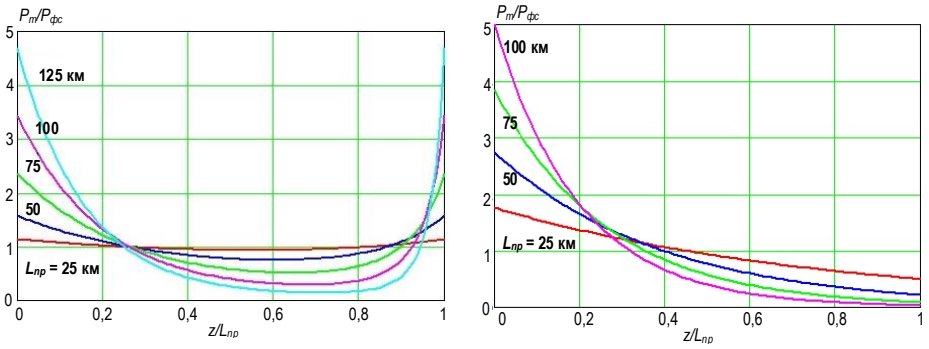
На рисунке 5а представлены зависимости нормированной относительной ПМ $p_n(z) = P_m(z)/P_{Фс}$ от относительного расстояния $z/L_{пр}$ в пределах одного пролета для встречной накачки. Для сравнения с методом поддержания квазисолитонного режима применением дискретных EDFA на рисунке 5b показаны зависимости, аналогичные показанным на рисунке 5а. Были определены по выражениям (6): среднее значение относительной ПМ в пролете p_0 и значение ПМ на входе пролета P_{0m}. Для расчетов использовалось значение мощности фундаментального солитона $P_{Фс} = 4{,}7$ мВт [2], которое было получено в результате моделирования ВОСС на том же ОМОВ, но без учета потерь. Результаты расчетов приведены в таблице 2.

ТАБЛИЦА 2. Результаты расчетов по выражению (6)

TABLE 2. Results of Calculations by Expression (6)

$L_{пр}$, км	25	50	75	100	125
p_0	0,881	0,636	0,424	0,290	0,214
P_{0m}, мВт	5,38	7,40	11,09	16,19	22,0

На рисунке 6 приведена схема квазисолитонной ВОСС с двумя пролетами. В исследованиях использовались схемы и с бóльшим количеством пролетов (до 20). УУ на схеме разделены оптическими изоляторами.



a) b)

Рис. 5. Зависимости нормированной относительной мощности $p_n(z)$ от относительного расстояния $z/L_{пр}$ в пределах одного УУ: для RA с встречной накачкой (a); дискретных EDFA (b)

Fig. 5. The Dependences of the Normalized Relative Power p(z) on the Relative Distance $z/L_{пр}$ within One Amplifying Section: a) для RA с встречной накачкой; b) для дискретных EDFA

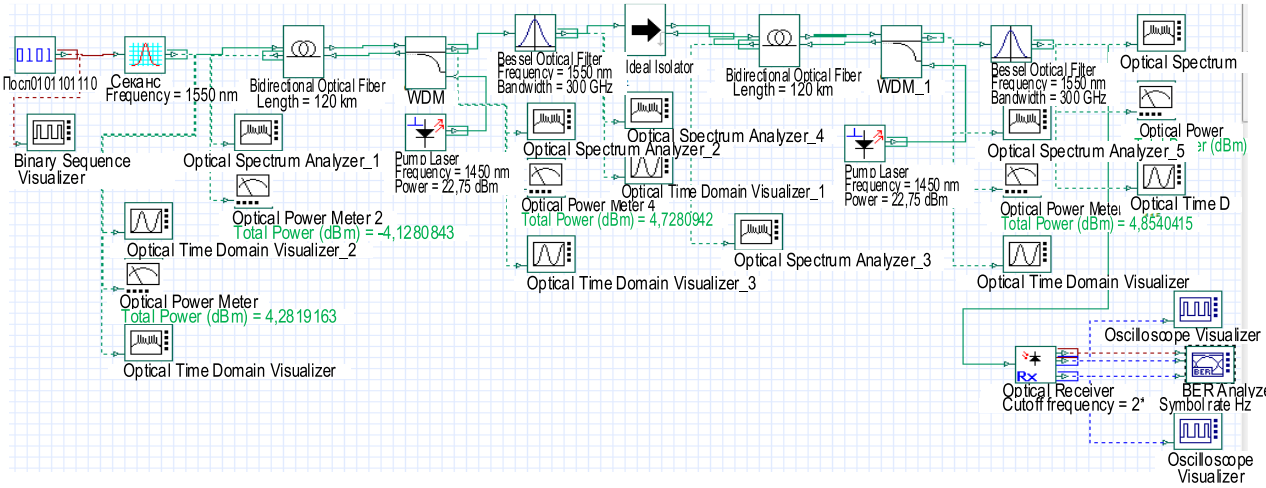
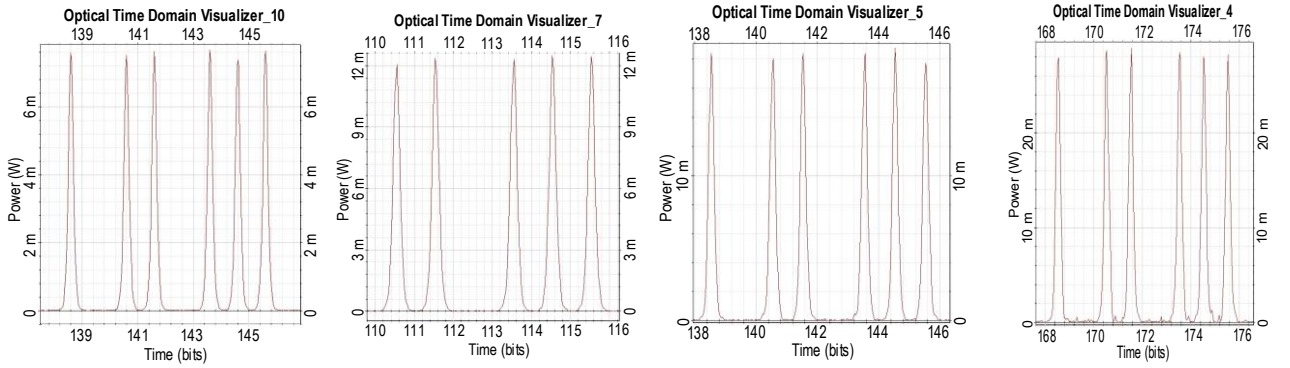


Рис. 6. Схема из двух последовательно соединенных УУ

Fig. 6. A Diagram of two series-connected amplifying sections

Для учета эффекта Рамана в схеме использовались более сложные двунаправленные модели ОМОВ, учитывающие процессы распространения излучения сигнала и накачки в ОМОВ, как вперед, так и назад. Мощность встречной накачки P_p подбиралась в процессе исследования по ПМ выходного импульса P_{lm}, которая должна равняться входной $P_{lm} = P_{0m}$ (таблица 2).

При моделировании длина пролета варьировалась от 25 до 125 км, а их количество от 1 до 20. Общая длина линии достигала 525 км. Результаты исследований приведены на рисунке 7 и в таблице 3.



a) b) c) d)

Рис. 7. Форма импульсов на выходах многопролетных квазисолитонных ВОСС: длина 50 км × 10 = 500 км (a); 75 км × 7 = 525 км (b); 100 км ×5 = 500 км (c); 125 км × 4 = 500 км (d)

Fig. 7. The Shape of Pulses at the Outputs of Multi-Span Quasi-Soliton Fiber-Optic Communication Systems: length 50 km × 10 = 500 km (a); 75 km × 7 = 525 km (b); 100 km ×5 = 500 km (c); 125 km × 4 = 500 km (d)

ТАБЛИЦА 3. Результаты расчетов и имитационного моделирования для квазисолитонной ВОСС с RA и встречной накачкой

TABLE 3. Results of Calculations and Measurements for a Fiber-Optic Communication System with Counter-Pumping of Raman

Установленные и расчетные величины							Результаты имитационного моделирования		
Количество пролетов	Длина ОМОВ, км	Мощность лазера накачки, дБм	Коэф-т усиления ОУ, дБ	Расчетная входная ПМ, мВт	Длительность входного импульса, пс	Установка входной ПМ, мВт	Выходная ПМ, мВт	Длительность выходного импульса, пс	*Q* фактор
1	25	17,29	5	5,38	20	5,8	5,8	20	309
2	50	17,29	5	5,38	20	5,8	5,8	20	234
4	100	17,29	5	5,38	20	5,8	5,8	20	176
8	200	17,29	5	5,38	20	5,8	5,7–5,9	19,8–20	123
12	300	17,29	5	5,38	20	5,8	5,7–6	19,7–19,9	107
20	500	17,29	5	5,38	20	5,8	5,4–5,8	19,6–20	70
1	50	19,25	10	7,40	20	7,8	7,85	20	266
2	100	19,25	10	7,40	20	7,8	7,8	20	169
4	200	19,25	10	7,40	20	7,8	7,8–8	19,8–20,1	126
8	400	19,25	10	7,40	20	7,8	7,6–7,9	19,9–20,4	80
10	500	19,25	10	7,40	20	7,8	7,5–7,9	19,7–20,3	65
1	75	20,76	15	11,09	20	12,5	12,6	20	236
2	150	20,76	15	11,09	20	12,5	12,7	19,8	148
4	300	20,76	15	11,09	20	12,5	12,7–13	19,8–20	100
6	450	20,76	15	11,09	20	12,5	12–12,6	19,6–20,1	71
7	525	20,76	15	11,09	20	12,5	11,8–12,5	19,7–20,5	66
1	100	21,96	20	16,19	20	18	18	20,5	186
2	200	21,96	20	16,19	20	18	18	20,6	123
3	300	21,96	20	16,19	20	18	18–18,6	19,9–20,5	82
4	400	21,96	20	16,19	20	18	17,3–18,4	20,8–21,3	65
5	500	21,96	20	16,19	20	18	17,9–18,5	20,1–20,9	54
1	125	22,95	25	22	20	25	25–25,5	21,7–22	124
2	250	22,95	25	22	20	25	25–27	22–22,8	77
3	375	22,95	25	22	20	25	25–27	20,5–23,2	42
4	500	22,95	25	22	20	25	27–29	20,1–21,2	45

Из таблицы 3 и рисунке 7 следует, что при длинах пролета от 0,25 до 1,25 L_D и общей длине ВОЛТ до 525 км квазисолитонный режим при встречной накачке сохраняется. Значения длительности секансного импульса и его амплитуда при длинах пролета до L_D сохраняются в пределах 2,5 и 4 %, соответственно. Однако при общей длине более 200 км наблюдается джиттер, который приводит к колебаниям длительности и амплитуды от импульса к импульсу в пределах 4 и 6 %, соответственно.

Многопролетные ВОСС с RA и двунаправленной накачкой

Рассмотрим применение распределенных ОУ ВКР с накачкой (рисунок 8). В этом случае для усиления оптических импульсов непрерывное излучение накачки подается с двух сторон пролета. Обозначения на схеме и исходные данные для расчета и моделирования такие же, как для ВОСС с встречной накачкой.

Выражения (2) и (3) справедливы и для пролета ВОСС с двунаправленной накачкой. Для изменений $g(z)$ в пролете вместо (4) можно записать в виде:

$$g(\mathrm{z}) = g_1 \cdot \exp(-\alpha_p z) + g_2 \exp[-\alpha_p (L_{\text{пр}} - z)], \quad (7)$$

где g_1 и g_2 связаны с мощностями попутной и встречной накачек.

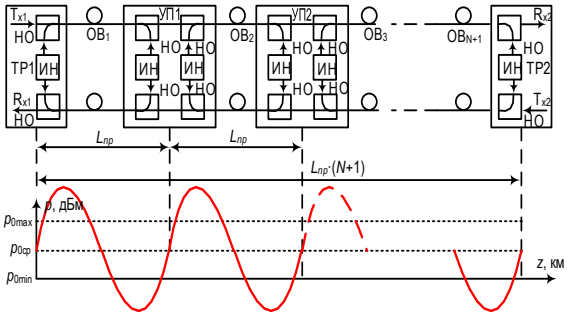


Рис. 8. Схема линейного тракта ВОСС с RA и двунаправленной накачкой

Fig. 8. Diagram of the Linear Path of a Fiber-Optic Communication System with Bidirectional Raman Pumping

Считая, что мощности встречной и попутной накачки равны, и решив уравнение (2) с учетом уравнения (7), получим зависимость мощности секансных импульсов в пролете для двунаправленной накачки [1–5]. Как и для встречной накачки, определим среднее значение нормализованной ПМ в пролете p_0 и значение ПМ на входе пролета P_{0m} с помощью выражения (8) для ВОСС с RA и двунаправленной накачки (таблица 4). На рисунке 9 представлены зависимости нормированной относительной ПМ $p_n(z) = P_m(z)/P_{\text{Фс}}$ от относительного расстояния $z/L_{\text{пр}}$ в пределах одного пролета для двунаправленной накачки.

$$P_m(z) = P_{0m} \cdot \exp\left\{\alpha L_{\text{пр}} \left(\frac{\text{sh}\left[\alpha_p \left(z - \frac{L_{\text{пр}}}{2}\right)\right] + \text{sh}\left(\frac{\alpha_p L_{\text{пр}}}{2}\right)}{2 \cdot \text{sh}\left(\frac{\alpha_p L_{\text{пр}}}{2}\right)} \right) - \alpha z \right\} = P_{0m} \cdot p(\mathrm{z}). \quad (8)$$

ТАБЛИЦА 4. Результаты расчетов по выражению (8)

TABLE 4. Results of calculations by expression (8)

$L_{\text{пр}}$, км	25	50	75	100	125
p_0	1	1,003	1,027	1,106	1,285
$P_{0\text{m}}$, мВт	4,7	4,69	4,58	4,25	3,66

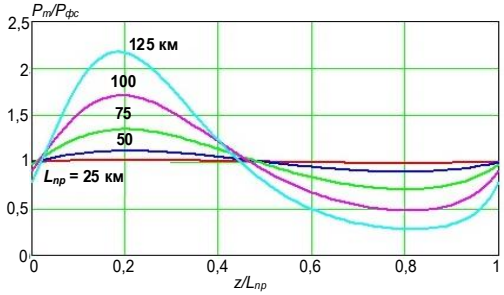


Рис. 9. Зависимости нормированной относительной мощности $p(z)$ от относительного расстояния $z/L_{\text{пр}}$ в пределах одного УУ

Fig. 9. The Dependences of the Normalized Relative Power p(z) on the Relative Distance $z/L_{\text{пр}}$ within One Amplifying Section

Как видно из таблицы 4, при использовании двунаправленной накачки ПМ входных импульсов P_{0m} при увеличении длины пролета становится даже меньше мощности $P_{\text{Фс}}$, необходимой для формирования фундаментального солитона.

Схема квазисолитонной ВОСС с двумя пролетами представлена на рисунке 10. При моделировании длина пролета варьировалась от 25 до 125 км, а их количество от 1 до 20. Общая длина ВОЛТ достигала 525 км. Результаты исследования приведены в таблице 5. В ней представлены результаты с использованием оптического фильтра Бесселя при длине УУ ≤ 50 км и без фильтра при длине пролета больше 50 км.

Из таблицы 5 и рисунка 11 следует, что при длинах пролета от 0,25 до 1,25 L_D и общей длине ВОЛС до 525 км квазисолитонный режим при двунаправленной накачке сохраняется. Значения длительности секансного импульса и его амплитуда при изменении количества пролетов имеют незначительные изменения. Однако при длине пролета 1,25 L_D наблюдается джиттер, который приводит к колебаниям длительности и амплитуды от импульса к импульсу в пределах 10 и 7,5 %, соответственно.

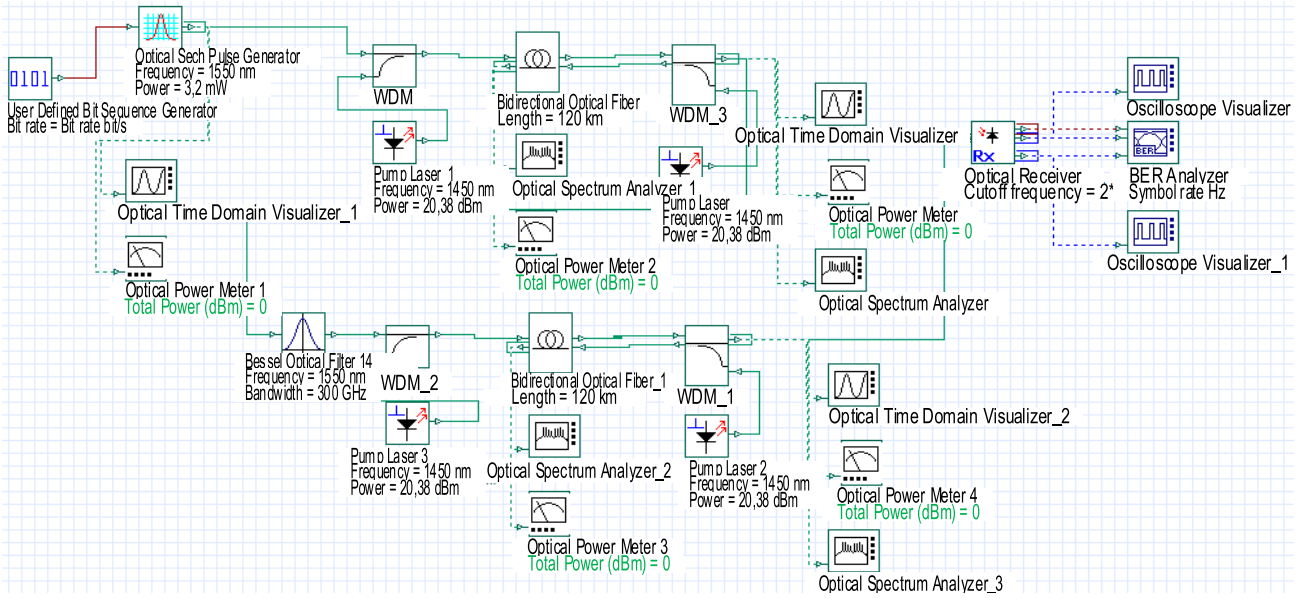
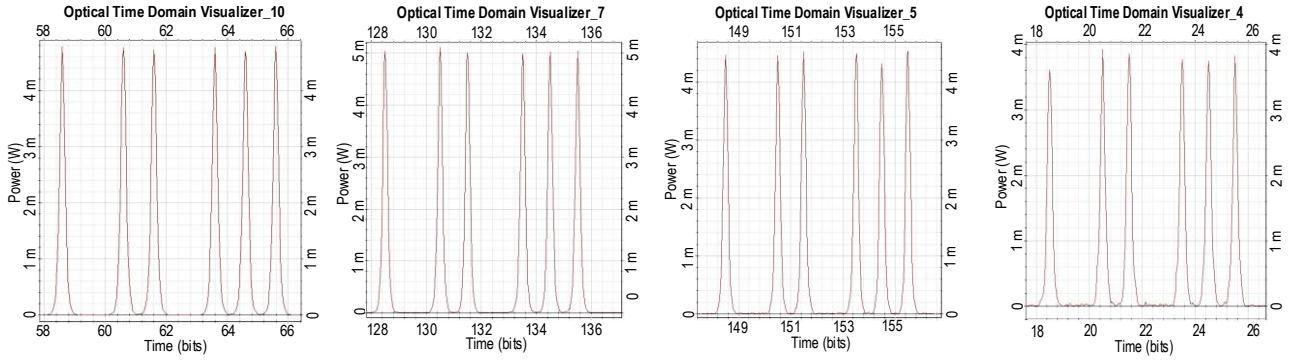


Рис. 10. Схема из двух последовательно соединенных пролетов

Fig. 10. A diagram of Two Series-Connected Amplifying Sections



a) b) c) d)

Рис. 11. Форма импульсов на выходах многопролетных квазисолитонных ВОСС: длина 50 км × 10 = 500 км (a); 75 км × 7 = 525 км (b); 100 км ×5 = 500 км (c); 125 км × 4 = 500 км (d)

Fig. 11. The Shape of Pulses at the Outputs of Multi-Span Quasi-Soliton Fiber-Optic Communication Systems: length 50 km × 10 = 500 km (a); 75 km × 7 = 525 km (b); 100 km ×5 = 500 km (c); 125 km × 4 = 500 km (d)

ТАБЛИЦА 5. Результаты расчетов и имитационного моделирования для ВОСС с RA и двунаправленной накачкой

TABLE 5. Results of Calculations and Measurements for a Fiber-Optic Communication System with Bidirectional Raman Pumping

Установленные и расчетные величины							Результаты имитационного моделирования		
Количество пролетов	Длина ОМОВ, км	Мощность лазеров накачки, дБм	Коэф-т усиления ОУ, дБ	Расчетная входная ПМ, мВт	Длительность входного импульса, пс	Установка входной ПМ, мВт	Выходная ПМ, мВт	Длительность выходного импульса, пс	*Q* фактор
1	25	14,31	5	4,7	20	4,7	4,7	20	304
2	50	14,31	5	4,7	20	4,7	4,65	20,3	249
4	100	14,31	5	4,7	20	4,7	4,65–4,7	20–20,1	200
8	200	14,31	5	4,7	20	4,7	4,6–4,7	20–20,3	129
12	300	14,31	5	4,7	20	4,7	4,7–4,9	19,5–20	96
20	500	14,31	5	4,7	20	4,7	4,7–5	19,6–20,4	65
1	50	16,24	10	4,69	20	4,7	4,75	20	241
2	100	16,24	10	4,69	20	4,7	4,7	20,1	191
4	200	16,24	10	4,69	20	4,7	4,65–4,8	20–20,2	133
8	400	16,24	10	4,69	20	4,7	4,7–4,8	19,9–20,2	82

Установленные и расчетные величины							Результаты имитационного моделирования		
Количество пролетов	Длина ОМОВ, км	Мощность лазеров накачки, дБм	Коэф-т усиления ОУ, дБ	Расчетная входная ПМ, мВт	Длительность входного импульса, пс	Установка входной ПМ, мВт	Выходная ПМ, мВт	Длительность выходного импульса, пс	*Q* фактор
10	500	16,24	10	4,69	20	4,7	4,7–4,9	19,8–20,1	69
1	75	17,71	15	4,58	20	4,7	4,7	20	200
2	150	17,71	15	4,58	20	4,7	4,7	20	159
4	300	17,71	15	4,58	20	4,7	4,75	20	98
6	450	17,71	15	4,58	20	4,7	4,7–4,9	19,7–20	77
7	525	17,71	15	4,58	20	4,7	5–5,1	19,6–20	61
1	100	18,92	20	4,25	20	4,5	4,6	20	162
2	200	18,92	20	4,25	20	4,5	4,5–4,6	19,7–20	126
3	300	18,92	20	4,25	20	4,5	4,5–4,7	19,7–20,1	92
4	400	18,92	20	4,25	20	4,5	4,4–4,6	19,8–20,6	76
5	500	18,92	20	4,25	20	4,5	4,3–4,6	19,6–20,5	63
1	125	19,88	25	3,66	20	4	4,1–4,2	20	150
2	250	19,88	25	3,66	20	4	4,0–4,1	20	89
3	375	19,88	25	3,66	20	4	3,8–4,0	20–21	70
4	500	19,88	25	3,66	20	4	3,7–4,0	20–22	56

Выводы

В статье исследованы способы поддержания квазисолитонного режима в многопролетных одноканальных ВОСС с дискретными ОУ, а также с RA со встречной и двунаправленной накачками. Все способы могут быть использованы в высокоскоростных ВОСС. Наиболее эффективным способом поддержания длительности и ПМ квазисолитонных импульсов на больших расстояниях является использование RA с двунаправленной накачкой. Это решение позволяет обеспечить максимальную длину пролета до 125 км при минимальных сигналах на входе пролетов с ПМ от 4 до 4,7 мВт практически равной или даже меньше, чем ПМ, необходимая для формирования фундаментальных солитонов в ОМОВ без потерь. Изменение длительности выходного импульса происходит в пределах от 19,5 до 22 пс. При использовании дискретных ОУ EDFA изменение длительности выходного импульса происходит в пределах от 18 до 26 пс, на входе пролетов использовались сигналы с ПМ от 7,5 до 22 мВт, значительно превосходящей ПМ, необходимую для формирования фундаментальных солитонов в ОМОВ без потерь. В таком случае необходимо использовать пролеты меньшей длины до 75 км и сокращать общую протяженность квазисолитонной ВОСС.

Промежуточное положение между рассмотренными способами поддержания квазисолитонного режима в многопролетных ВОСС занимает использование RA со встречной накачкой. Длительность выходного импульса изменяется в пределах от 19,6 до 23,2 пс, на входе пролетов использовались сигналы с ПМ от 5.8 до 25 мВт, значительно превосходящей ПМ необходимую для формирования фундаментальных солитонов в ОМОВ без потерь. В такой ВОЛТ пролеты не должны превышать длину 100 км.

Решение с использованием на ВОСС RA со встречной накачкой является более экономичным по сравнению с двунаправленной накачкой, т. к. позволяет использовать меньшее количество усилителей на каждом пролете, но значительно уступает в качестве связи. Значение *Q* фактора RA с двунаправленной накачкой в среднем на 5 % больше, чем при встречной накачке на длинах пролетов от 25 до 125 км и на 10 % больше, чем при дискретных ОУ EDFA с длиной пролета бóльшей или равной 75 км. При построении ВОСС, где требуется максимальная длина пролетов и повышенное качество связи, т. е. большие значения *Q* фактора, целесообразно использовать RA с двунаправленной накачкой.

Разработанные схемы квазисолитонных ВОСС и методики их исследования могут быть использованы в учебном процессе для подготовки специалистов по волоконно-оптической связи.

Список источников

1. Андреева Е.И., Былина М.С., Глаголев С.Ф., Доценко С.Э., Чаймарданов П.А. Свойства временных оптических солитонов в оптических волокнах и возможность их использования в телекоммуникациях. Часть 4 // Труды учебных заведений связи. 2019. Т. 5. № 1. С. 15–24. DOI:10.31854/1813-324X-2019-5-1-15-24

2. Глаголев С.Ф., Доценко С. Э. Поддержание квазисолитонного режима в ВОСС с использованием усилителей Рамана // XI Международная научно-техническая и научно-методическая конференция «Актуальные проблемы инфотелекоммуникаций в науке и образовании» (Санкт-Петербург, Россия, 15–16 февраля 2022). СПб: СПбГУТ, 2022. С. 343–348.
3. Агравал Г. Нелинейная волоконная оптика. М.: Мир, 1996. 323 с.
4. Agrawal G.P. Fiber-Optic Communication Systems. Wiley, 2012. 626 p.
5. Кившарь Ю.С., Агравал Г.П. Оптические солитоны. От волоконных световодов до фотонных кристаллов. М.: Физматлит, 2005. 648 с.
6. Листвин А.В., Листвин В.Н., Швырков Д.В. Оптические волокна для линий связи. М.: ЛЕСАРарт, 2003. 288 с.
7. OptiSystem User Guide and Reference Manual. Optical Communication System Design Software. Version 19. Optiwave Systems Inc. 2022.
8. Андреев В.А. Рамановские усилители на волоконно-оптических линиях передачи. М.: Ириас, 2008. 219 с.
9. Листвин В.Н., Трещиков В.Н. DWDM системы. М.: Техносфера, 2021. 420 с.
10. Леонов А.В., Наний О.Е., Трещиков В.Н. Усилители на основе вынужденного комбинационного рассеяния в оптических системах связи // Прикладная фотоника. 2014. Т. 1. № 1. С. 27–50.

References

1. Andreeva E, Bylina M., Glagolev S., Dotsenko S., Chaimardanov P. Properties of Temporary Optical Solitons in Optical Fibers and the Possibility of their Use in Telecommunications. Part 4. *Proceedings of Telecommunication Universities*. 2019;5(1):15–24. (in Russ.) DOI:10.31854/1813-324X-2019-5-1-15-24
2. Glagolev S., Dotsenko S. Maintaining a Quasi-Soliton Mode in Fiber-Optic Communication System Using Raman Amplifiers. *Proceedings of the* XI*th International Conference on Infotelecommunications in Science and Education, 15–16 February 2022, St. Petersburg, Russia*. St. Petersburg: The Bonch-Bruevich Saint-Petersburg State University of Telecommunications Publ.; 2022. p.343–348. (in Russ.)
3. Agraval G. *Nonlinear Fiber Optics*. Moscow: Mir Publ.; 1996. 323 p. (in Russ.)
4. Agrawal G.P. *Fiber-Optic Communication Systems*. Wiley; 2012. 626 p.
5. Kishvar Yu.S., Agraval G.P. *Optical Solitons. From Fiber Light Guides to Photonic Crystals.* Moscow: Fizmatlit Publ.; 2005. 648 p (in Russ.)
6. Listvin A.V., Listvin V.N., Shvyrkov D.V. *Optical Fibers for Communication Lines.* Moscow: LESARart Publ.; 2003. 288 p. (in Russ.)
7. *OptiSystem User Guide and Reference Manual. Optical Communication System Design Software*. Version 19. Optiwave Systems Inc. 2022.
8. Andreev V.A. *Raman Amplifiers on Fiber-Optic Transmission Lines*. Moscow: Irias Publ.; 2008. 219 p (in Russ.)
9. Listvin V.N., Treschikov N. *DWDM Systems.* Moscow: Technosphere Publ; 2021. 420 p. (in Russ.)
10. Leonov A.V., Nanii O.E., Treschikov V.N. Raman Amplifiers in Optical Communication Systems. Applied Photonics. 2014;1(1):27–50. (in Russ.)

Статья поступила в редакцию 03.03.2023; одобрена после рецензирования 01.04.2023; принята к публикации 21.04.2023.
The article was submitted 03.03.2023; approved after reviewing 01.04.2023; accepted for publication 21.04.2023.

Информация об авторах:

ГЛАГОЛЕВ Сергей Федорович — кандидат технических наук, доцент кафедры фотоники и линий связи Санкт-Петербургского государственного университета телекоммуникаций им. проф. М.А. Бонч-Бруевича
https://orcid.org/0000-0002-0664-9877

ДОЦЕНКО Сергей Эдуардович — инженер кафедры фотоники и линий связи Санкт-Петербургского государственного университета телекоммуникаций им. проф. М.А. Бонч-Бруевича
https://orcid.org/0000-0003-0299-0469

Научная статья
УДК 681.518.5
DOI:10.31854/1813-324X-2023-9-2-57-64

Итерационное совмещение геометрически подобных изображений с использованием контуров

Ринат Радмирович Диязитдинов, rinat.diyazitdinov@gmail.com

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

Аннотация: *Совмещение геометрически подобных изображений проводится по методике с раздельной оценкой параметров. В декартовой системе координат оценивается смещение вдоль координатных осей, в логарифмически-полярной системе – оцениваются масштаб и поворот. Для повышения точности оценки параметров модели (смещений, масштаба и поворота) обработка данных проводится итерационным способом. Для уменьшения времени совмещения предлагается вместо сравнения изображений по коэффициенту корреляции использовать сравнение контуров по количеству совпадающих точек. Для проверки разработанной методики использовались кадры с изображением вагона. Выигрыш по времени обработки модифицированной методики «со сравнением контуров» оценивался в сравнении с исходной методикой «со сравнением изображений».*

Ключевые слова: *совмещение, контур, изображение, итерационный, декартовый, логарифмически-полярный, время обработки*

Ссылка для цитирования: Диязитдинов Р.Р. Итерационное совмещение геометрически подобных изображений с использованием контуров // Труды учебных заведений связи. 2023. Т. 9. № 2. С. 57–64. DOI:10.31854/1813-324X-2023-9-2-57-64

Superposition of the Similarity Images by Contour

Rinat Diyazitdinov, rinat.diyazitdinov@gmail.com

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

Abstract: *Superposition of the similarity images is implemented by the methodology with divided estimation of parameters. The offsets along the coordinate axes are estimated in the Cartesian coordinate system. The scale and the rotate are estimated in the log-polar coordinate system. The accurate estimation of parameters of similarity models (offsets, scale and rotate) is achieved by the iteration processing. Optimization of the processing time is achieved by contour comparison instead of the image comparison. The test data for experiment is image with the freight car. The decreasing of the processing time for the modified methodology of "contour comparison" was estimated by comparison with the source methodology of "image comparison".*

Keywords: *superposition, contour, image, Iteration, Cartesian, log-polar, processing time*

For citation: Diyazitdinov R. Superposition of the Similarity Images by Contour. *Proc. of Telecom. Universities.* 2023;9(2):57–64. (in Russ.) DOI:10.31854/1813-324X-2023-9-2-57-64

Введение

Задача совмещения изображений является актуальной задачей в интеллектуальных системах видеонаблюдения.

Совмещение используется при решении таких практических задач, как:

– формирование общего изображения из изображений-фрагментов (формирование панорамных фотографий);

– сравнение изображений в медицинских целях (определение изменений и патологий в органах человека);

– поиск кадра в видеопотоке по заданному видеоизображению;

– измерение расстояний и трехмерная реконструкция в результате совмещения изображений, полученных от различных камер;

– повышение эффективности работы оператора за счет совмещения изображений низкого и высокого разрешения (просмотр изображений на экране одного монитора с автоматическим переключением между данными разных камер при масштабировании эффективнее, чем просмотр двух мониторов с изображениями различных масштабов) и т. д.

Для совмещения изображений широко применяется модель геометрического подобия, которая, с одной стороны, достаточно просто описывается (модель содержит четыре параметра), а с другой стороны, достаточно точно описывает реальные преобразования [1, 2].

Обработка изображений требует огромных вычислительных затрат, что связано с огромным объемом данных, содержащихся в них. В данной статье предлагается алгоритм совмещения геометрически подобных изображений, который позволяет сократить время обработки за счет использования контуров.

Модель

Модель геометрического подобия, связывающая совмещаемые изображения, описывается четырьмя параметрами: двумя смещениями вдоль координатных осей, масштабом и поворотом:

$$\begin{aligned} x' &= x\cdot\alpha\cdot\cos(\varphi) - y\cdot\alpha\cdot\sin(\varphi) + h; \\ y' &= y\cdot\alpha\cdot\sin(\varphi) + y\cdot\alpha\cdot\cos(\varphi) + p, \end{aligned} \quad (1)$$

где (x, y), (x', y') – координаты точек в системах координат, связанных моделью подобия; h, p – смещения, α – масштаб, φ – угол поворота.

Изображения $f(x_i, y_i)$ и $g(x_i, y_i)$, связанные моделью геометрического подобия, описываются следующими формулами:

$$\begin{aligned} f(x_i, y_i) &= s(x_i, y_i) + k(x_i, y_i); \\ g(x_i, y_i) &= \lambda s(x'_i, y'_i) + \gamma + m(x_i, y_i), \end{aligned} \quad (2)$$

где $s(x_i, y_i)$ – детерминированное, но неизвестное изображение; $k(x_i, y_i)$, $m(x_i, y_i)$ – помехи; λ, γ – мультипликативная и аддитивная составляющая, соответственно.

Совмещение геометрически подобных изображений является актуальной задачей, которую можно рассматривать как в качестве самостоятельной задачи, так и в качестве задачи, решаемой в составе комплексных. Типичным примером комплексных задач является оценка параметров аффинного преобразования и преобразование перспективы, в которых идея оценки базируется на предварительном сопоставлении «небольших» фрагментов по модели подобия, после чего проводится непосредственная оценка параметров интересуемой модели преобразования.

Примечание: логика подобного подхода заключается в том, что для «небольших» фрагментов отличия между интересуемой и моделью подобия незначительны [1, 2].

Однако одной из основных проблем при оценке параметров модели подобия является значительное время обработки. Например, время обработки изображений 720×1080 пикселей может достигать нескольких десятков минут в зависимости от используемого алгоритма, размера совмещаемого фрагмента и условий съемки. По этой причине актуальной проблемой в задаче совмещения геометрически подобных изображений является сокращение времени обработки.

Решением этой проблемы занимались различные исследователи. Ниже представлены работы, в которых приведены приемы для уменьшения времени совмещения изображений.

Релевантные работы

В работе [3] приведен алгоритм совмещения изображений, связанных аффинным преобразованием. Алгоритм предполагает выполнение следующих этапов: вычисление контура на изображении, определение особых точек на контуре и сопоставление особых точек по дескрипторам. На основе сопоставленных точек вычисляются параметры совмещения. Недостатком этой работы является узкая область применения – обработка медицинских изображений, на которых фактически содержится единственный контур, описывающий человеческий орган.

Авторы работ [4, 5] предлагают методику проективного совмещения изображений, которая также основана на совмещении контуров. Но вместо сопоставления особых точек по дескрипторам используется процедура проверки гипотез, которые определяются четырьмя точками контура. Оценка параметров совмещения определяется той гипотезой, которая обеспечивает максимальное совпадение контуров. Недостатком данных работ

является специфическая обработка, связанная с выделением контура, которая применима только для изображения поверхности Земли.

В исследованиях [1, 2] представлен алгоритм совмещения изображений, который использует особые точки изображений. Сопоставление точек определяется с помощью дескрипторов, а параметры совмещения рассчитываются по методу наименьших квадратов. Недостатком работы, по заявлению самих авторов, является 5-процентная вероятность неверного сопоставления. Для устранения этого недостатка могут быть использованы способы обработки, обеспечивающие борьбу с импульсными выбросами/неверным сопоставлениями [6], однако их использование существенно увеличивает время обработки.

Работа [7] предлагает алгоритм совмещения, который очень похож на алгоритм из работ [1, 2], но сопоставление особых точек проводится по методу «ближайшего соседа». Параметры аффинного преобразования для совмещения изображений определяются с помощью метода наименьших квадратов. Недостатки данной работы такие же, как у предыдущей.

Как можно видеть, основное направление для уменьшения времени совмещения изображений – это использование особых точек. Однако необходимость проверки правильности сопоставления или узкая область применения является сдерживающим фактором, который не позволяет разработать одновременно и «универсальную» и «быструю» методику совмещения изображений.

В основе данной работы лежит методика, описанная в статье [8]. Однако вместо сравнения изображений по коэффициенту корреляции будет использоваться сравнение предварительно распознанных контуров по количеству совпадающих контурных точек. Это отличие позволит значительно уменьшить время обработки, так как сравнение контуров проводится за гораздо меньшее время, чем сравнение изображений.

Методика

Методика совмещения, описанная в работе [8], рассматривает следующую ситуацию: первое изображение представляет собой фрагмент второго. Для совмещения (для оценки смещений, масштаба и поворота) необходимо точке первого изображения поставить в соответствие точку второго изображения (примечание: в таком случае эти точки будут являться реперными). Это позволит оценить масштаб и поворот с помощью корреляционно-экстремального способа [9, 10], используя представление изображений в логарифмически-полярной системе координат. Оценив масштаб и поворот, и зная координаты реперных точек, оцениваются параметры смещений.

Таким образом, если известны реперные точки на изображениях, то задача совмещения изображений имеет достаточно простое решение. Однако определение реперных точек на изображениях может быть трудоемкой задачей. Например, пусть первое изображение имеет размеры 100×100 пикселей, а второе – 320×240 пикселей. В качестве реперной точки на первом изображении может выступать центральная точка. А чтобы определить для этой точки соответствие на втором изображении, нужно проверить все точки второго изображения, то есть $320 \cdot 240 = 76\,800$ точек. Такое большое количество проверок требует огромных вычислительных затрат, и, как следствие, значительного времени обработки.

Для уменьшения времени обработки в работе [8] была представлена методика итерационной обработки. Для совмещения достаточно, чтобы точка, выбранная на втором изображении, находилась в некоторой окрестности относительно реперной точки (рисунок 1). Методика итерационной обработки позволяет совмещать изображения при таком выборе точек на изображениях, даже если они не являются реперными. Эта особенность позволяет выбирать точки с некоторым шагом. Например, если шаг равен 5 пикселям, то количество точек, которое нужно проверить на втором изображении уменьшится в 5^2 раз, то есть $320 \cdot 240/5^2 = 3\,072$ точки (рисунок 2).

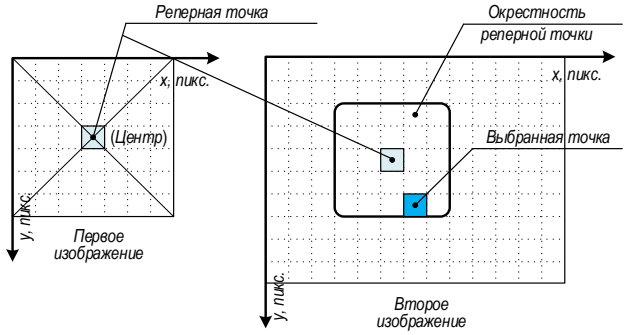


Рис. 1. Выбор точки на изображении для совмещения

Fig. 1. Searching the Point for Image Superposition

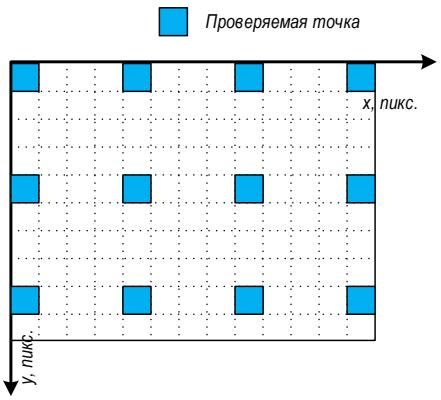


Рис. 2. Проверяемые точки на изображении для методики итерационной обработки

Fig. 2. The Points of Image for Iteration Processing

Таким образом, методика итерационной обработки позволяет значительно уменьшить время совмещения изображений.

Методика итерационной обработки состоит из следующих шагов.

Шаг 1. Загрузка изображений.

Шаг 2. Определение реперной точки на первом изображении. Она будет соответствовать центральной точке – (X_0, Y_0).

Шаг 3. Определение совокупности выбранных точек на втором изображении (u_i, w_i), i – 1 .. K, где K – количество выбранных точек.

Шаг 4. Проверка i-ой выбранной точки (U_0, W_0) = = (u_i, w_i), которая выявляет параметры совмещения и коэффициент корреляции, определяющий меру совпадения обрабатываемых изображений (3), где $f(x_i, y_i|\ \theta)$ – первое изображение (фрагмент), преобразованное в соответствии с параметрами совмещения θ; $g(x_i, y_i)$ – второе изображение; N – количество пикселей (элементов) изображения; θ = {h, p, φ, α, λ, γ} – параметры совмещения.

$$R(\theta) = \frac{(\sum_{i=1}^{N} g(x_i,y_i)\cdot f(x_i,y_i|\theta))/N - ((\sum_{i=1}^{N} g(x_i,y_i))/N)\cdot((\sum_{i=1}^{N} f(x_i,y_i|\theta))/N)}{((\sum_{i=1}^{N} g^2(x_i,y_i))/N - (\sum_{i=1}^{N} g(x_i,y_i)/N)^2)^{1/2}\cdot\left((\sum_{i=1}^{N} f^2(x_i,y_i|\theta))/N - \left(\sum_{i=1}^{N} f((x_i,y_i|\theta))/N\right)^2\right)^{1/2}}, \quad (3)$$

Шаг 4 соответствует итерационной обработке, которая выполняется поэтапно.

Этап 1. Определение исходных точек (X_0, Y_0), (U_0, W_0) на первом и втором изображении, соответственно.

Этап 2. Расчет положения точек в соответствии с номером итерации j:

– если $j = 1$, то $X_0^j = X_0$, $Y_0^j = Y_0$, $U_0^j = U_0$, $W_0^j = W_0$;

– если $j > 1$, то $X_0^j = X_0$, $Y_0^j = Y_0$.

$$U_0^j = X_0 \cdot \alpha^{j-1}\cos(\varphi^{j-1}) - Y_0 \cdot \alpha^{j-1}\sin(\varphi^{j-1}) + h^{j-1}, \quad (4)$$

$$W_0^j = X_0 \cdot \alpha^{j-1}\sin(\varphi^{j-1}) + Y_0 \cdot \alpha^{j-1}\cos(\varphi^{j-1}) + p^{j-1}, \quad (5)$$

где h^{j-1}, p^{j-1}, α^{j-1}, φ^{j-1} – смещения, масштаб, угол поворота, оцененные на $(j-1)$-й итерации.

Этап 3. Проверка на завершение итерационной обработки.

Если:

$$\left|U_0^j - U_0^{j-1}\right| < TU \text{ и } \left|W_0^j - W_0^{j-1}\right| < TW, \quad (6)$$

где TU, TW – это пороги (примечание: для проведения экспериментов значения порогов выбирались равными 0,5 пикселя), то итерационная процедура завершается, в противном случае проводится уточнение параметров.

Этап 4. Преобразование первого изображения в логарифмически-полярную систему координат относительно точки (X_0^j, Y_0^j).

Этап 5. Преобразование второго изображения в логарифмически-полярную систему координат относительно точки (U_0^j, W_0^j).

Этап 6. Оценка масштаба и поворота по изображениям в логарифмически-полярной системе координат [8, 11] – α^j, φ^j.

Этап 7. Преобразование первого изображения в соответствии с оцененным масштабом и поворотом. Масштабирование и поворот изображения проводится относительно точки (X_0^j, Y_0^j). Это эквивалентно трем преобразованиям:

– смещение изображения, чтобы точка (X_0^j, Y_0^j) совпала с точкой (0, 0):

$$M1 = \begin{bmatrix} 1 & 0 & -X_0^j \\ 0 & 1 & -Y_0^j \\ 0 & 0 & 1 \end{bmatrix},$$

– масштабирование и поворот в соответствии с α^j, φ^j:

$$M2 = \begin{bmatrix} \alpha^j\cos(\varphi^j) & -\alpha^j\sin(\varphi^j) & 0 \\ \alpha^j\sin(\varphi^j) & \alpha^j\cos(\varphi^j) & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

– смещение изображения для возврата в точку (X_0^j, Y_0^j):

$$M3 = \begin{bmatrix} 1 & 0 & X_0^j \\ 0 & 1 & Y_0^j \\ 0 & 0 & 1 \end{bmatrix}.$$

Преобразование, эквивалентное этим трем последовательным преобразованиям, записывается следующим образом:

$$M = M3 \cdot M2 \cdot M1 = \begin{bmatrix} \alpha^j\cos(\varphi^j) & -\alpha^j\sin(\varphi^j) & X_0^j - X_0^j\cdot\alpha^j\cos(\varphi^j) + Y_0^j\cdot\alpha^j\sin(\varphi^j) \\ \alpha^j\sin(\varphi^j) & \alpha^j\cos(\varphi^j) & Y_0^j - X_0^j\cdot\alpha^j\sin(\varphi^j) - Y_0^j\cdot\alpha^j\cos(\varphi^j) \\ 0 & 0 & 1 \end{bmatrix} \quad (7)$$

Этап 8. Оценка смещений по преобразованному первому (этап 7) и второму изображению в декартовой системе координат [10] – *H, P.*

Этап 9. Вычисление смещений h^j, p^j:

$$h^j = X_0^j - X_0^j \cdot \alpha^j \cos(\varphi^j) + Y_0^j \cdot \alpha^j \sin(\varphi^j) + H, \quad (8)$$

$$p^j = Y_0^j - X_0^j \cdot \alpha^j \sin(\varphi^j) - Y_0^j \cdot \alpha^j \cos(\varphi^j) + P. \quad (9)$$

Этап 10. Преобразование первого изображения в соответствии с оцененным масштабом, поворотом и смещениям согласно матрице:

$$\begin{bmatrix} \alpha^j \cos(\varphi^j) & -\alpha^j \sin(\varphi^j) & h^j \\ \alpha^j \sin(\varphi^j) & \alpha^j \cos(\varphi^j) & p^j \\ 0 & 0 & 1 \end{bmatrix}. \quad (10)$$

Этап 11. Оценивание параметров $\{\lambda^j, \gamma^j\}$ по методу наименьших квадратов:

$$\lambda^j = \frac{Sg \cdot Sf - N \cdot Sfg}{Sf^2 - N \cdot Ef}, \gamma^j = \frac{Sg \cdot Sfg - Sg \cdot Ef}{Sf^2 - N \cdot Ef}, \quad (11)$$

где

$$Sf = \sum_{i=1}^{N} f(x_i, y_i|\theta), Sg = \sum_{i=1}^{N} g(x_i),$$

$$Sfg = \sum_{i=1}^{N} f(x_i, y_i|\theta) \cdot g(x_i), Ef = \sum_{i=1}^{N} f^2(x_i, y_i|\theta),$$

$$\theta = \{h^j, p^j, \varphi^j, \alpha^j\},$$

$f(x_i, y_i|\theta)$ – сигнал после преобразования с учетом найденный смещений и угла поворота.

Этап 12. Вычисление коэффициента корреляции по формуле (3).

Этап 13. Переход к этапу 2.

В описанной методике большую часть процессорного времени занимает оценка масштаба, поворота и смещений (этапы 6 и 8), так как в них происходит сравнение изображений множество раз (рисунок 3).

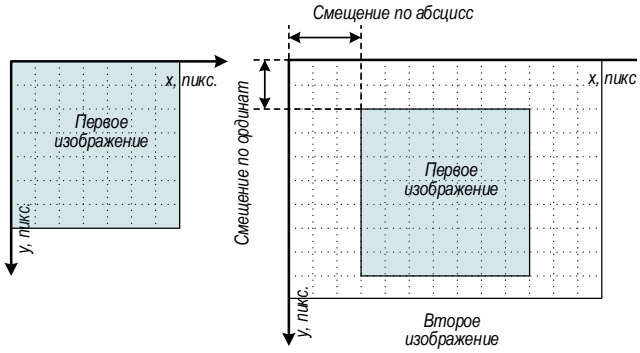


Рис. 3. Процесс сравнения изображений

Fig. 3. Comparison of Images

Шаг 5. Из *K* вариантов совмещений выбирается тот, который соответствует максимальному коэффициенту корреляции, то есть такому совмещению, при котором изображения будут совпадать наилучшим образом.

Количество сравнений будет равно количеству проверяемых смещений, а каждое сравнение – это расчет коэффициента корреляции (примечание: корреляция используется в качестве метрики сходства изображений [9–11]). Операция расчета коэффициента корреляции является очень затратной с точки зрения процессорного времени (примечание: для ускорения вычислений можно использовать преобразование Фурье [12, 13], однако использование этого преобразования не приводит к кардиальному уменьшению времени обработки).

Для уменьшения времени совмещения была предложена модификация методики, которая представлена в следующем параграфе.

Модификация методики совмещения

Для уменьшения времени обработки предлагается распознавать на изображениях контуры и использовать их для оценки параметров совмещения. В контурах количество элементов (контурных точек) намного меньше, чем количество элементов (пикселей) на изображениях. Вместо вычисления корреляции (этапы 6 и 8) будет вычисляться количество совпавших точек между контурами (примечание: количество совпавших точек используется в качестве метрики сходства контуров; точки считаются совпавшими, если расстояние между ними меньше некоторого порога [3, 4, 14]). Такой подход гораздо выгоднее с точки зрения времени, затрачиваемого на обработку.

Таким образом, за счет уменьшения количества обрабатываемых элементов (контурных точек вместо пикселей) и за счет более быстрого вычисления метрики «сходства» (совпадение контурных точек вместо коэффициента корреляции) предложенная модификация будет обеспечивать более высокую скорость совмещения изображений.

В следующем параграфе представлен эксперимент, показывающий результаты обработки в соответствии с исходной и модифицированной методикой.

Эксперимент

Для определения времени обработки в соответствии с исходной и модифицированной методикой совмещения был проведен эксперимент с использованием реальных кадров с изображением вагона.

На рисунке 4 показаны примеры совмещаемых изображений.

В результате совмещения изображений (рисунок 5) были оценены параметры совмещения:

$$\hat{\alpha} = 0{,}6598; \hat{\varphi} = 30°; \hat{h} = 200{,}4 \text{ пикселей};$$
$$\hat{p} = 21{,}0 \text{ пикселей}, \hat{\lambda} = 1{,}1808, \hat{\gamma} = 21{,}8487.$$

Время совмещения составило 2 минуты.

Рис. 4. Совмещаемые изображения

Fig. 4. The Images for Superposition

Рис. 5. Совмещение изображений

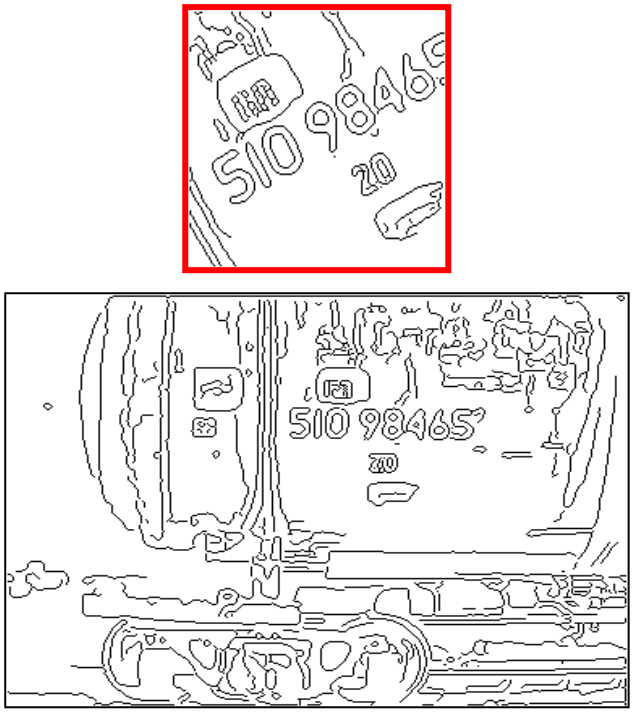
Fig. 5. The Result of Images Superposition

Усовершенствованная методика с предварительным распознаванием контуров по алгоритму Canny [15] (рисунок 6) оценила параметры совмещения как:

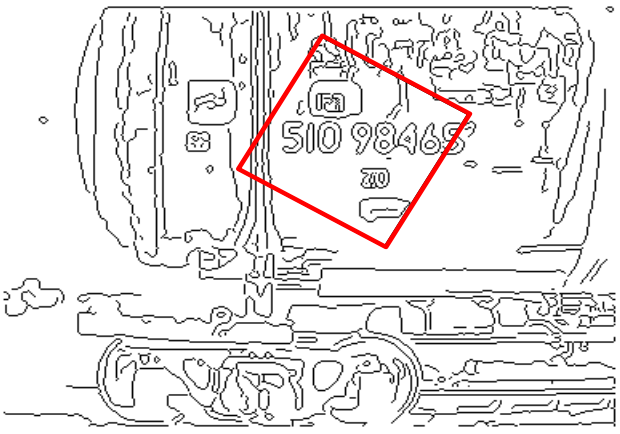
$$\hat{\alpha} = 0{,}6598;\ \hat{\varphi} = 30°;\ \hat{h} = 200{,}0 \text{ пикселей};$$

$$\hat{p} = 21{,}2 \text{ пикселей};\ \hat{\lambda} = 1{,}1803;\ \hat{\gamma} = 21{,}7039.$$

Время совмещения составило 20 секунд.



a)



b)

c)

Рис. 6. Распознанные контуры (a), совмещение контуров (b), совмещение изображений по параметрам, оцененным при совмещении контуров (c)

Fig. 6. Detecting Contours (a), the Result of Contours Superposition (b), the Result of Images Superposition by Parameters of Contours Superposition (c)

Как можно видеть, результаты совмещения (см. рисунки 5 и 6с) практически одинаковые, однако время обработки сократилось в 6 раз. Таким образом, предложенная модификация позволяет значительно уменьшить время обработки для совмещения изображений.

Заключение

В работе представлена модификация методики итерационного совмещения изображений. Модификация заключается в замене сравнения изображений по коэффициенту корреляции на сравнение контуров по количеству совпадающих точек.

Обработка контура отличается от обработки изображения тем, что в контуре меньше количества элементов (контурных точек), чем элементов (пикселей) на изображении, а процедура подсчета совпадающих точек занимает меньше времени, чем расчет коэффициента корреляции. За счет этого предложенная модификация позволила значительно уменьшить время совмещения изображений. Скорость обработки в описанном эксперименте возросла в 6 раз.

С точки зрения качества совмещения модифицированная методика не ухудшила результаты обработки по сравнению с исходной методикой. Визуальных различий между результатами обработки не наблюдается.

Представленная методика совмещения изображений с использованием контуров может быть применена для совмещения изображений в интеллектуальных системах видеорегистрации. Например, для поиска кадров в видеопотоке, совмещения фрагментов изображений, принадлежащих различным кадрам, в целях проведения измерений [16], совмещения разномасштабных изображений, чтобы повысить разрешающую способность, и т. д.

В подобных интеллектуальных системах одной из важнейших технико-эксплуатационных характеристик является время совмещения. Представленная модификация оптимизирует этот параметр, что говорит о целесообразности ее применения и внедрения в интеллектуальные системы видеорегистрации.

Список источников

1. Ke Y., Sukthankar R. PCA-SIFT: A more distinctive representation for local image descriptors // Computer Vision and Pattern Recognition. 2004. Vol. 2. PP. 506–513. DOI:10.1109/CVPR.2004.1315206
2. Bay H., Tuytelaars T., Van Gool L. Surf: Speeded up robust features // Proceedings of the 9th European Conference on Computer Vision (ECCV 2006, Graz, Austria, 7–13 May 2006). Lecture Notes in Computer Science. Vol. 3951. Berlin, Heidelberg: Springer, 2006. PP. 404–417. DOI:10.1007/11744023_32
3. Сунгатуллина Д., Крылов А. Быстрый алгоритм совмещения контуров изображений, связанных изотропным аффинным преобразованием // GraphiCon. 2014. С. 92–95. URL: https://www.graphicon.ru/html/2014/papers/92-95.pdf (дата обращения 11.05.2023)
4. Новиков А.И., Саблина В.А., Горячев Е.О. Применение контурного анализа для совмещения изображений // Известия Тульского государственного университета. Технические науки. 2013. № 9-1. С. 260–270.
5. Елесина С.И., Ефимов А.И. Отбор критериальных функций для систем улучшенного и комбинированного видения // Известия Тульского государственного университета. Технические науки. 2013. № 9-1. С. 229–236.
6. Raguram R., Frahm J.M., Pollefeys M. A Comparative Analysis of RANSAC Techniques Leading to Adaptive Real-Time Random Sample Consensus // Proceedings of the 10th European Conference on Computer Vision (ECCV 2008, Marseille, France, 12–18 October 2008). Lecture Notes in Computer Science. Vol. 5303. Berlin, Heidelberg: Springer, 2008. PP. 500–513. DOI:10.1007/978-3-540-88688-4_37
7. Создание бесшовного изображения. URL: https://wiki.gis-lab.info/w/Создание_бесшовного_изображения (дата обращения 31.03.2023)
8. Диязитдинов Р.Р., Васин Н.Н. Итерационный алгоритм оценки смещения и угла поворота при влиянии аддитивной и мультипликативной помехи для пространственно-временного совмещения телевизионных сигналов // Труды учебных заведений связи. 2020. Т. 6. № 4. С. 28–34. DOI:10.31854/1813-324X-2020-6-4-28-34
9. Кузьмин С.В. Инвариантное к масштабу определение задержек между двумя одномерными цифровыми сигналами // Инфокоммуникационные технологии. 2011. Т. 9. № 2. С. 7–10.
10. Губанов А.В., Ефимов В.М., Киричук В.С., Пустовских А.И., Резник А.Л. Методы оценивания взаимного смещения фрагментов изображений // Автометрия. 1988. № 3. С. 70–73.
11. Мясников Е.В. Определение параметров геометрических трансформаций для совмещения портретных изображений // Компьютерная оптика. 2007. Т. 31. № 3. С. 77–82.
12. De Castro E., Morandi C. Registration of Translated and Rotated Images Using Finite Fourier Transforms // IEEE Transactions Pattern Analysis and Machine Intelligence. 1987. Vol. 9. Iss. 5. PP. 700–703. DOI:10.1109/TPAMI.1987.4767966
13. Reddy B.S., Chatterji B.N. An FFT-based technique for translation, rotation, and scale-invariant image registration // IEEE Transactions Pattern Analysis and Machine Intelligence. 1996. Vol. 5. Iss. 8. PP. 1266–1270. DOI:10.1109/83.506761
14. Ефимов А.И., Новиков А.И. Алгоритм поэтапного уточнения проективного преобразования для совмещения изображений // Компьютерная оптика. 2016. Т. 40. № 2. С. 258–265. DOI:10.18287/2412-6179-2016-40-2-258-265
15. Canny J.F. A Computational Approach To Edge Detection // IEEE Transactions on Pattern Analysis and Machine Intelligence. 1986. Vol. 8(6). PP. 679–698. DOI:10.1109/TPAMI.1986.4767851

16. Диязитдинов Р.Р., Васин Н.Н. Использование фрагментов телевизионного изображения системы технического зрения для верификации повышения помехоустойчивости измерений скорости протяженного объекта // Труды учебных заведений связи. 2022. Т. 8. № 1. С. 34–40. DOI:10.31854/1813-324X-2022-8-1-34-40

References

1. Ke Y., Sukthankar R. PCA-SIFT: A more distinctive representation for local image descriptors. *Computer Vision and Pattern Recognition*. 2004;2:506–513. DOI:10.1109/CVPR.2004.1315206

2. Bay H., Tuytelaars T., Van Gool L. Surf: Speeded up robust features. *Proceedings of the 9th European Conference on Computer Vision, ECCV 2006, 7–13 May 2006, Graz, Austria. Lecture Notes in Computer Science. vol.3951*. Berlin, Heidelberg: Springer; 2006. p.404–417. DOI:10.1007/11744023_32

3. Sungatullina D., Krilov A. A fast algorithm for contours image superposition, which connected by an isotropic affine transformation. *GraphiCon*. 2014:92–95. (in Russ.) URL: https://www.graphicon.ru/html/2014/papers/92-95.pdf [Accessed 11th May 2023]

4. Novikov A.I., Sablina V.A., Goryachev E.O. Contour analysis application for image superimposition. *Izvestiya TulGU. Technical Sciences*. 2013;9-1:260–270. (in Russ.)

5. Elesina S.I., Efimov A.I. Selection of criterial functions for combined and enhanced synthetic vision systems. *Izvestiya TulGU. Technical Sciences*. 2013;99-1:229–236. (in Russ.)

6. Raguram R., Frahm J.M., Pollefeys M. A Comparative Analysis of RANSAC Techniques Leading to Adaptive Real-Time Random Sample Consensus. *Proceedings of the 10th European Conference on Computer Vision, ECCV 2008, 12–18 October 2008, Marseille, France. Lecture Notes in Computer Science, vol.5303*. Berlin, Heidelberg: Springer; 2008. p.500–513. DOI:10.1007/978-3-540-88688-4_37

7. Creating a Seamless Image. (in Russ.) URL: https://wiki.gis-lab.info/w/Создание_бесшовного_изображения [Accessed 31st March 2023]

8. Diyazitdinov R., Vasin N. Iterative Algorithm Offset and Angle Rotation Estimation with Additive and Multiplicative Noise for Space-Time Superposition of Television Signals. *Proc. of Telecom. Universities*. 2020;6(4):28–34 (in Russ.) DOI:10.31854/1813-324X-2020-6-4-28-34

9. Kuzmin S.V. Scale-invariant delay estimation between two one-dimensional digital signals. *Infokommunikacionnye tehnologii*. 2011;9(2):7–10. (in Russ.)

10. Gubanov A.V., Efimov V.M., Kirichuk V.S., Pustovskih A.I., Reznik A.L. Methods for offset estimating of image fragments. *Avtometriya*. 1988;3:70–73. (in Russ.)

11. Myasnikov E.V. Geometric Transform Parametrs Estimation for Superposition Portrait Images. *Computer Optics*. 2007;31(3):77–82. (in Russ.)

12. De Castro E., Morandi C. Registration of Translated and Rotated Images Using Finite Fourier Transforms. *IEEE Transactions Pattern Analysis and Machine Intelligence*. 1987;9(5):700–703. DOI:10.1109/TPAMI.1987.4767966

13. Reddy B.S., Chatterji B.N. An FFT-based technique for translation, rotation, and scale-invariant image registration. *IEEE Transactions Pattern Analysis and Machine Intelligence*. 1996;5(8):1266–1270. DOI:10.1109/83.506761

14. Efimov A.I., Novikov A.I. An Algorithm for Multistage Projective Transformation Adjustment for Image Superimposition. *Computer Optics*. 2016;40(2):258–265. (in Russ.) DOI:10.18287/2412-6179-2016-40-2-258-265

15. Canny J.F. A Computational Approach To Edge Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 1986;8(6):679–698. DOI:10.1109/TPAMI.1986.4767851

16. Diyazitdinov R., Vasin N. Using Television Image Fragments of a Machine Vision for Verifying Noise Immunity of an Extended Object Velocity Measurement. *Proc. of Telecom. Universities*. 2022;8(1):37–40. (in Russ.) DOI:10.31854/1813-324X-2022-8-1-37-40

Статья поступила в редакцию 01.04.2023; одобрена после рецензирования 03.04.2023; принята к публикации 10.04.2023.
The article was submitted 01.04.2023; approved after reviewing 03.04.2023; accepted for publication 10.04.2023.

Информация об авторе:

ДИЯЗИТДИНОВ Ринат Радмирович
кандидат технических наук, доцент, доцент кафедры Сетей и систем связи Поволжского государственного университета телекоммуникаций и информатики
https://orcid.org/0000-0001-6360-0351

Научная статья
УДК 621.391
DOI:10.31854/1813-324X-2023-9-2-65-71

Анализ характеристик алгоритмов прекодирования сигналов в системе MU-MIMO с группированием абонентов

Александр Александрович Калачиков, 330rts@gmail.com

Сибирский государственный университет телекоммуникаций и информатики,
Новосибирск, 630102, Российская Федерация

Аннотация: *В статье представлены результаты имитационного моделирования алгоритма прекодирования ZF с использованием алгоритма ортогонального выбора абонентов и алгоритма максимизации взаимной информации в нисходящей системе MU-MIMO. При количестве пользователей бóльшем, чем количество антенн на базовой станции, возникает взаимная корреляция между каналами пользователей, что снижает суммарную спектральную эффективность системы MU-MIMO. Для снижения эффекта взаимной корреляции применяется подбор пользователей на основе максимальной ортогональности между ними. Суммарная спектральная эффективность в системе MU-MIMO зависит от условий реального распространения сигналов и для каналов с пространственной корреляцией необходимо использовать выбор подмножества абонентов с низкой корреляцией между их векторами каналов. Для исследования эффективности выбора подмножества абонентов используется модель канала с открытым исходным кодом, позволяющая получать реалистичные реализации канала.*

Ключевые слова: *5G new radio, QuaDRIGa, 3GPPмодель канала, прекодирование ZF, система MU-MIMO, группирование абонентов*

Ссылка для цитирования: Калачиков А.А. Анализ характеристик алгоритмов прекодирования сигналов в системе MU-MIMO с группированием абонентов // Труды учебных заведений связи. 2023. Т. 9. № 2. С. 65–71. DOI:10.31854/1813-324X-2023-9-2-65-71

Numerical Evaluation of the MU-MIMO Beamforming Performance with User Selection Algorithm

Александр Калачиков, 330rts@gmail.com

Siberian State University of Telecommunications and Informatics,
Novosibirsk, 630102, Russian Federation

Abstract: *This paper presents the numerical evaluation of the ZF beamforming algorithm using the user selection in the multiuser multiantenna (MU-MIMO) downlink system. Two user selection algorithm – semiorthogonal user selection and greedy user selection are numerically evaluated based on the open source MIMO channel model. The sum rate performance depending on number of users are presented. The arising inter user correlation degrades the sum rate (spectral efficiency) performance of multiuser MIMO system especially in scenarios where the number of users is larger than the number of antennas at the BS. The selection of users is based on the orthogonality of the channels among selected users. For MIMO channel simulation the QUADRIGA channel model reflecting the real propagation conditions is used. The obtained performance of MU-MIMO ZF precoding in spatially correlated channel are compared based on the empirical cumulative density function of the sum rate of multiple users. Numerical results*

show that the ZF precoder using user selection (G ZF) outperforms the ZF precoder with random user selection in spectral efficiency. The greedy user selection in spatially correlated channel has advantage to semi-orthogonal user selection. It isobserved that as the increasing the number of served users used for selection the greedy user selection gives better performance than semi-orthogonal algorithm.

Keywords: *5G new radio, QuaDRIGa, 3GPP channel model, ZF precoding, multiuser MIMO system, user selection*

For citation: Kalachikov A. Numerical Evaluation of the MU-MIMO Beamforming Performance with User Selection Algorithm. *Proc. of Telecom. Universities.* 2023;9(2):65–71. (in Russ.) DOI:10.31854/1813-324X-2023-9-2-65-71

I. Введение

Технология многоантенных систем связи MIMO (*аббр. от англ.* Multiple Input, Multiple Output) является базовой в современных мобильных сетях для повышения спектральной эффективности (СЭ) системы связи за счет перегрузки (повторного использования) часототно-временных ресурсов системы, одновременной передачи сигналов нескольких пользователей на этих ресурсах при соответствующем прекодировании сигналов. Прекодирование состоит в обработке независимых сигналов пользователей при формировании передаваемого сигнала многоантенной системой. При обработке происходит умножение передаваемых символов на комплексные весовые векторы, что определяет пространственные свойства передаваемых сигналов. СЭ прекодирования зависит от конкретных условий распространения сигналов в радиоканале между базовой станцией (БС) и абонентами. В практических сценариях развертывания сети абоненты распределены случайно, но во многих случаях, особенно в городской застройке, расположены с большей плотностью. Это приводит к коррелированности векторов каналов абонентов и снижению эффективности прекодирования, уменьшается СЭ при количестве пользователей большем, чем количество антенн БС [1].

В системах MU-MIMO рассматриваются подоптимальные методы линейного прекодирования, позволяющие получить выигрыш от пространственного мультиплексирования. К ним относятся, в частности, алгоритм обнуления интерференции (ZF, *аббр. от англ.* Zero Forcing), основанный на обращении матрицы канала пользователей, метод блочной диагонализации (BD, *аббр. от англ.* Block Diagonalization), основанный на вычислении сингулярного разложения матрицы канала пользователей, регуляризированные варианты указанных алгоритмов, позволяющие улучшить их характеристики в каналах с наличием пространственной корреляции [2].

Если количество пользователей превышает количество антенн БС, используются алгоритмы выбора подмножества пользователей, для которых применяется прекодирование и повышается суммарная СЭ. Выбор основан на принципе ортогональности векторов каналов между выбранными пользователями. Каналы в подмножестве должны иметь низкую коррелированность, быть почти ортогональными. Пользователи с высокой корреляцией между каналами выбираются алгоритмом распределения ресурсов для передачи на другом частотно-временном блоке [2]. Изучение числовых характеристик прекодирования с использованием методов выбора в сценариях практического развертывания необходимо с точки зрения практической реализации как часть задачи комплексного моделирования алгоритмов на системном уровне.

В данной статье представлены результаты численного моделирования алгоритмов прекодирования с использованием выбора подмножества абонентов на модели канала, реалистично отражающей распространение сигналов в заданном сценарии. В качестве алгоритма прекодирования используется алгоритм ZF, достаточно просто реализуемый стандартными вычислительными функциями.

II. Теория

A. Модель системы связи

На стороне БС системы связи используется антенна из N_T элементов, обслуживающая K пользователей, каждый из которых оборудован одной антенной. Коэффициенты передачи канала описываются вектором коэффициентов $\mathbf{h}_k \in C^{N_T x 1}$. Набор индексов активных обслуживаемых абонентов $U \subset \{1 \dots K\}$, набор индексов выбранных для прекодирования абонентов $S \subset U$.

Вектор передаваемых символов $\mathbf{x}_k \in C^{N_T x 1}$ составляется в следующем виде:

$$\mathbf{x}_k = \sum_{k=1}^{K} \mathbf{w}_k \, s_k,$$

где s_k – передаваемый символ данных пользователя k; $\mathbf{w}_k \in C^{N_T x 1}$ – вектор прекодирования пользователя k.

Принятый вектор сигналов пользователя k на поднесущей с номером s и номером символа n запишется в виде:

$$y_{k,n,s} = \mathbf{h}_{k,n,s}^T \mathbf{x}_{n,s} + n_{k,n,s} \text{ для } k = 1, \dots, K. \quad (1)$$

где $n_{k,n,s}$ – комплексный гауссовский шум с нулевым матожиданием и дисперсией σ_k^2.

В матричной форме матрица канала MU-MIMO составляется из векторов каналов пользователей $\mathbf{H}_{n,s} = [\mathbf{h}_{1,n,s} \dots \mathbf{h}_{K,n,s}]^T$ и вектор принятых сигналов пользователей определяется выражением:

$$\mathbf{y}_{n,s} = \mathbf{H}_{n,s}^T \mathbf{x}_{n,s} + n_{n,s}. \quad (2)$$

В системе 5G NR оценивание канала выполняется при использовании опорных сигналов на пилотных поднесущих в составе OFDM-сигнала (*аббр. от англ.* Orthogonal Frequency-Division Multiplexing – мультиплексирование с ортогональным частотным разделением каналов). Оценивание канала связи на стороне БС выполняется в частотной области на основе пилот-сигналов SRS (*аббр. от англ.* Sounding Reference Signals). Каждому абоненту присваивается определенная последовательность, размещенная по полосе частот системы. После приема пилот-сигналов SRS на стороне БС вычисляется частотная характеристика канала, на поднесущих с данными пользователей частотная характеристика интерполируется. Различные конфигурации опорных сигналов в восходящем направлении приведены в [3]. Сигналы SRS передаются в составе OFDM-символов с определенным расположением по поднесущим и номерам слотов. Для каждой последовательности SRS настраиваются количество антенных портов, символы в слоте, соответствующие каждой последовательности, слоты в периоде передачи, плотность пилотных поднесущих в полосе частот сигнала.

Последовательность SRS генерируется путем циклического сдвига базисной последовательности, которая формируется из последовательности Задова-Чу. Базовая последовательность определяется как $r_{u,v}^{SRS} = r_{u,v}^{\alpha}(n), n = 0,..,M_{SC}^{RS}$, где M_{SC}^{RS} – длина последовательности опорного сигнала; $u = 0,..,29$ – номер группы базисной последовательности; $v = 0,1$ – номер последовательности в группе. Базовая последовательность циклически сдвигается для увеличения общего числа доступных последовательностей. Сигналы SRS от различных пользователей могут быть мультиплексированы по частоте в пределах полосы частот сигнала, используя различные комбинированные шаблоны в соответствии с частотным смещением [3].

Оценивание канала выполняется в частотной области с использованием алгоритма наименьших квадратов (LS, *аббр. от англ.* Least Squares). На пилотных поднесущих SRS оцениваются коэффициенты частотной характеристики делением принятого значения сигнала на значение элемента SRS последовательности для каждого абонента. Значения коэффициентов передачи на поднесущих с данными пользователя интерполируется линейной или кубической интерполяцией.

При использовании прекодирования принятый сигнал для пользователя k определяется в виде:

$$y_k = \mathbf{h}_k^T \mathbf{w}_k s_k + \sum_{j \neq k} \mathbf{h}_k^T \mathbf{w}_j s_j + \mathbf{n}_k, \quad (3)$$

где сумма соответствует сигналам интерференции со стороны других пользователей [4].

Суммарная СЭ вычисляется как сумма по всем пользователям и поднесущим и зависит от соотношения сигнал/(шум + интерференция) SINR (*аббр. от англ.* Signal Interference + Noise Ratio) каждого пользователя.

Величина SINR пользователя k на одной поднесущей вычисляется по выражению:

$$\mathrm{SINR}_k = \frac{|\mathbf{h}_k^T \mathbf{w}_k|^2}{\sum_{j \neq k} |\mathbf{h}_k^T \mathbf{w}_j|^2 + K\sigma^2/P}, \quad (4)$$

где в числителе находится мощность сигнала пользователя k, в знаменателе сумма мощности переданных сигналов других пользователей (сигналы интерференции) и внутренних тепловых шумов [4].

Достижимая суммарная СЭ по всем пользователям вычисляется в виде $R_{BF} = \sum_{k=1}^{K} (\log_2(1 + \mathrm{SINR}_k))$, бит/с/Гц. Данная величина используется как метрика, показатель качества, при моделировании алгоритмов прекодирования.

В. Частично ортогональный выбор пользователей

Алгоритмы выбора абонентов являются частью системы распределения ресурсов и планирования нагрузки системы связи. Данные алгоритмы выполняют поиск группы абонентов K в виде подмножества всех обслуживаемых абонентов $\widehat{K}$ данной БС. Соответствующая комбинация пользователей K находится по критерию максимизации суммарной СЭ – R_{BF}.

Субоптимальный метод выбора абонентов с ограниченной ортогональностью (SUS, *аббр. от англ.* Semi orthogonal User group Selection) выполняет последовательный подбор абонентов с допущением на неполную ортогональность между векторами каналов [5]. На первой итерации алгоритма выбирается набор параметров – набор выбранных пользователей $\mathbf{S}^0 = \mathbf{0}$, набор оставшихся пользователей $T^{\{1\}} = \{1 \dots K\}$.

Используя метод ортогонализации Грамма-Шмидта, на n-й итерации вычисляется компонента $\mathbf{g}_k$ вектора канала $\mathbf{h}_k$, ортогональная подпространству, сформированная векторами каналов предыдущих выбранных пользователей $\mathbf{g}_1 \dots \mathbf{g}_{n-1}$:

$$\mathbf{g}_k = \mathbf{h}_k - \sum_{j=1}^{i-1} \frac{\mathbf{h}_k \mathbf{g}_j^*}{||\mathbf{g}_j||^2} \mathbf{g}_j. \quad (5)$$

В соответствии с вычисленным значением $\mathbf{g}_k$ n-й пользователь выбирается по наибольшей норме проекции $\|\mathbf{g}_k\|^2$ и вектор $\mathbf{h}_k$ удаляется из списка оставшихся пользователей. Из набора оставшихся пользователей выбираются почти ортогональные вектору $\mathbf{g}_k$, и процесс итерационно повторяется, пока набор оставшихся пользователей не окажется пустым $T^{\{n+1\}} = \mathbf{0}$ [5].

C. Выбор пользователей на основе алгоритма поглощения

Алгоритм поглощения (*от англ.* Greedy User Selection) выполняет подбор пользователей на основе вычисления СЭ каждого при максимизации суммарной СЭ [6]. На этапе инициализации для каждого пользователя с вектором канала $\mathbf{h}_k$ вычисляется коэффициент $r_k = \mathbf{h}_k^H \mathbf{h}_k$ и находится первый выбранный пользователь с индексом s_1: $s_1 = \text{argmax}(r_k), k \in U$. Устанавливается множество выбранных пользователей $S_1 = \{s_1\}$ и вычисляется СЭ – $R_{BF}(S_1)$; после это – суммарная СЭ по оставшимся пользователям. Находится пользователь s_k по выражению:

$$s_k = \text{argmax}\big(R_{BF}(S_{k-1} \cup s_1)\big).$$

Устанавливается $S_k = S_{k-1} \cup s_k$ и вычисляется $R_{BF} = R_{BF}(S_k)$. Если $R_{BF}(S_k) \le R_{BF}(S_{k-1})$, поиск заканчивается и выбирается $S = S_{k-1}$.

Алгоритм поглощения последовательно добавляет пользователей к набору выбранных, если суммарная СЭ при этом увеличивается. На первой итерации выбирается пользователь с максимальным значением нормы вектора канала и вычисляется СЭ; на последующих итерациях – их суммарная СЭ. Если она на n-й итерации снижается по сравнению с предыдущей (n – 1) итерацией, то n-й пользователь отбрасывается. Для выбранных пользователей вычисляются весовые вектора прекодирования по алгоритму обнуления сигналов интерференции ZF.

D. Прекодирование по алгоритму ZF

Алгоритм ZF вычисляет весовые векторы прекодирования для снижения взаимной интерференции между всеми пользователями данного частотно-временно́го ресурса. Алгоритм использует псевдообращение матрицы канала и на ее основе вычисляет векторы прекодирования. Вектор прекодирования ZF $\mathbf{w}_k$ пользователя k ортогонален комплексно сопряженным векторам каналов всех остальных пользователей $\mathbf{h}_k^H \mathbf{w}_j = 0$ для $j \ne k$. Матрица прекодирования $\mathbf{W}_{ZF}$ составляется из векторов прекодирования всех пользователей и вычисляется на основе матрицы канала $\mathbf{H}_{n,s}$ как ее псевдообращение $\mathbf{W}_{ZF} = \mathbf{H}(\mathbf{H}^H\mathbf{H})^{-1}$. Матрица канала $\mathbf{H}_{n,s}$ оценивается на пилотных поднесущих.

В реальных условиях распространения радиоканал является пространственно-коррелированным [7]. При этом элементы матрицы канала $\mathbf{H}_{n,s}$ взаимно коррелированны и это свойство значительно определяет характеристики системы MU-MIMO. Вычисление обратной матрицы коррелированного канала бывает затруднительно и матрица канала $\mathbf{H}_{n,s}$ становится плохо обусловленной, что приводит к увеличению уровня шума в пространственных каналах с большим значением собственных чисел [8].

E. Модель канала QuaDRiGa

Для решения задач моделирования и определения характеристик прекодирования MU-MIMO в практических сценариях развертывания используется модель канала 5GNR. Модель канала с открытым исходным кодом QuaDRiGa является 3GPP-3D геометрической вероятностной моделью, реализованной в соответствии с требованиями стандарта [9] с параметризацией по результатам многочисленных измерений радиоканалов.

Для проведения имитационного моделирования используется реализация модели канала 3GPP 3D с открытым исходным кодом QuaDRIGa [11, 12]. Модель позволяет получить реалистичные реализации канала при различных сценариях развертывания. Модель канала QuaDRiGa параметризируется в соответствии с необходимым сценарием развертывания сети. По выбранному сценарию генерируется расположение кластеров объектов, отражающих и рассеивающих колебания. Угловое рассеяние отраженных сигналов, параметры крупномасштабных и мелкомасштабных замираний генерируются вероятностным методом для каждой реализации канала.

Основные параметры системы моделирования определяются перед генерированием реализаций канала. Положение БС, конфигурация антенн, траектория перемещения абонентов и соответствующие сценарии распространения являются исходными параметрами. Генерация реализаций канала состоит из генерации параметров крупномасштабных замираний (рассеяние задержек, угловое рассеяние) и положения рассеивающих кластеров объектов [11, 12].

III. Результаты моделирования

В данном разделе представлены результаты моделирования алгоритма ZF с использованием двух алгоритмов выбора пользователей. Характеристики прекодирования оцениваются по величине средней СЭ с использованием реализаций канала, полученных в модели канала QuaDRIGa. Параметры моделирования приведены в таблице 1.

ТАБЛИЦА 1. Параметры моделирования

TABLE 1. Simulation Parameters

Параметр модели	Значение
Тип модели	QuaDRIGa v.2.2
Сценарий	3GPP 38.901 UMiNloS
Количество многолучевых компонент	12
Центральная частота	3,6 ГГц
Полоса	20 МГЦ
Количество поднесущих	400
Количество пользователей	4–36
Количество пользователей в группе	4
Количество антенн на БС	8–16

В соответствии с сценарием городской застройки и малой размерности зоны покрытия БС (UMi) пользователи распределены равномерно вокруг БС на удалении до 500 м от БС. Реализация канала для данного пользователя и элемента антенны на БС состоит из многолучевых компонент. Частотная характеристика каждого канала в полосе 20 МГц делится на 400 поднесущих передаваемого сигнала. Для каждой поднесущей вычисляется вектор прекодирования.

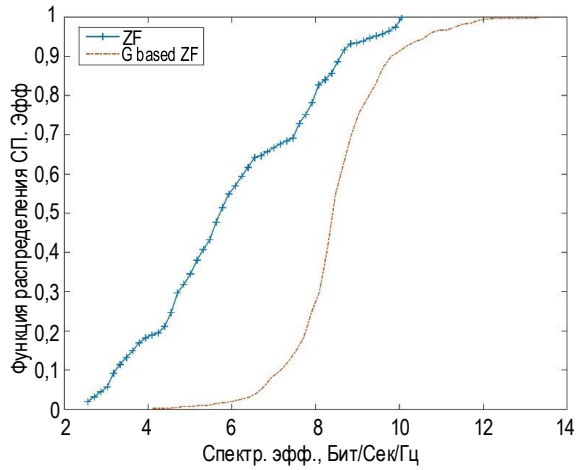
Канал каждого пользователя и весовые векторы прекодеров нормируются к единичной мощности. Мощность передатчика выбирается $P = 1$, мощность шума – $\sigma_n{}^2$, отношение сигнал/шум – $SNR = 1/\sigma^2 = 12$ дБ. Для сравнения эффективности прекодирования при различных условиях распространения были выбраны две величины числа обусловленности, соответствующие различной степени пространственной корреляции. Число обусловленности CN = 240 соответствует каналу со средней пространственной корреляцией, CN = 500 соответствует сильно коррелированному каналу.

Результаты моделирования показаны в виде функций распределения СЭ. На рисунке 1 показано распределение суммарной СЭ при прекодировании ZF для случайного выбора абонентов и для выбора по алгоритму SUS для подмножества из 4-х пользователей, выбираемых из общего множества 36 пользователей.

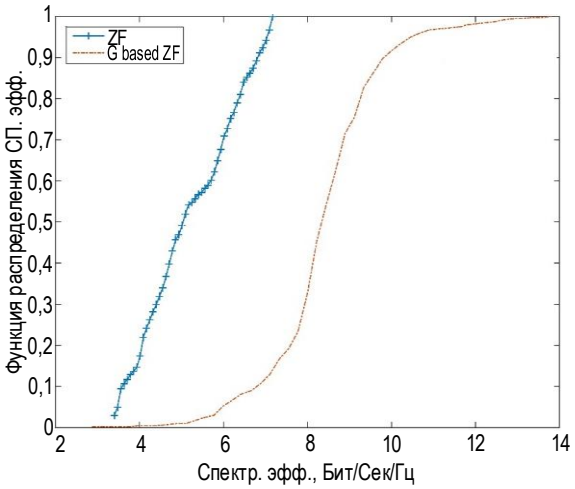
Для данных условий распространения прекодер ZF с выбором абонентов SUS показывает бóльшую СЭ по сравнению со случайным выбором абонентов. Прекодер с выбором по алгоритму SUS (график с меткой *G* ZF) достигает среднего значения СЭ 8 бит/сек/Гц, прекодер ZF со случайным выбором абонентов достигает среднего значения СЭ 5 (см. рисунок 1a) и 4,5 (см. рисунок 1b) бит/сек/Гц при 16 антеннах на БС. Число обусловленности канала для первого случая CN = 240, а для второго – CN = 500, т. е. при бóльшей пространственной корреляции.

Сравнение суммарной СЭ прекодирования ZF и выбора абонентов по методу SUS для различного числа пользователей, используемых для выбора, представлено на рисунке 2. Набор состоит из 4-х и 8-и пользователей. При увеличении их количества СЭ увеличивается, что показывает влияние алгоритма выбора абонентов на суммарную СЭ при количестве обслуживаемых пользователей большем, чем количество антенн на стороне БС.

На рисунке 3 приводится распределение суммарной СЭ в случае без прекодирования, при прекодировании ZF для случайного выбора абонентов, при прекодировании ZF с выбором по алгоритму SUS (SUS ZF) и выбором по алгоритму поглощения (Greedy ZF). Для подмножества 8-и пользователей, выбираемых из общего множества 40 пользователей. Число обусловленности канала CN = 140.



a)



b)

Рис. 1. Распределение СЭ при прекодировании ZF и выборе абонентов SUS, число обусловленности канала *CN* = 240 (a) и *CN* = 500 (b)

Fig. 1. Distribution of Sum Rate for ZF Precoding and User Selection SUS for Condition Number CN = 240 (a) and CN = 500 (b)

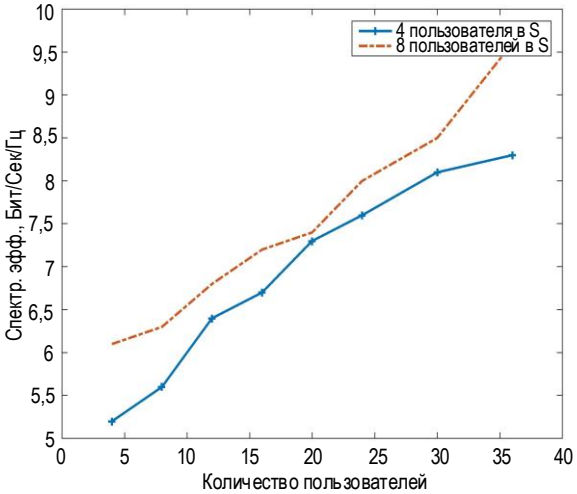


Рис. 2. Сравнение суммарной СЭ прекодирования ZF для различного числа пользователей

Fig. 2. Comparison of Sum Rate using ZF Precoding for Different Nimber of Users

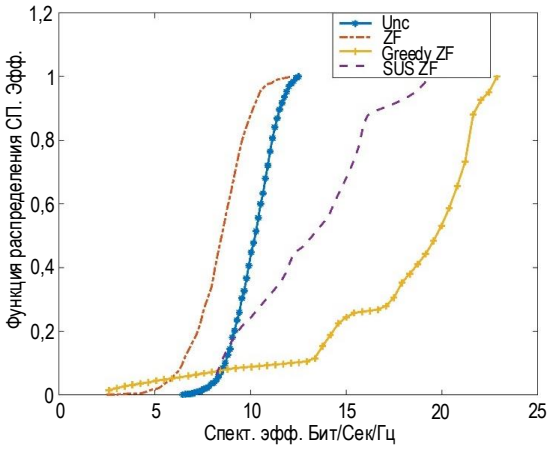


Рис. 3. Распределение СЭ при прекодировании ZF и выборе абонентов SUS, Greedy

Fig. 3. Distribution of Sum Rate for ZF Precoding and User Selection SUS, Greedy

В данных условиях распространения средняя СЭ системы без прекодирования оказалась выше СЭ прекодирования ZF и случайным выбором абонентов. Алгоритм поглощения показывает бóльшую СЭ по сравнению с алгоритмом SUS.

На рисунке 4 представлено сравнение полученной СЭ для двух способов выбора абонентов в зависимости от количества обслуживаемых БС абонентов.

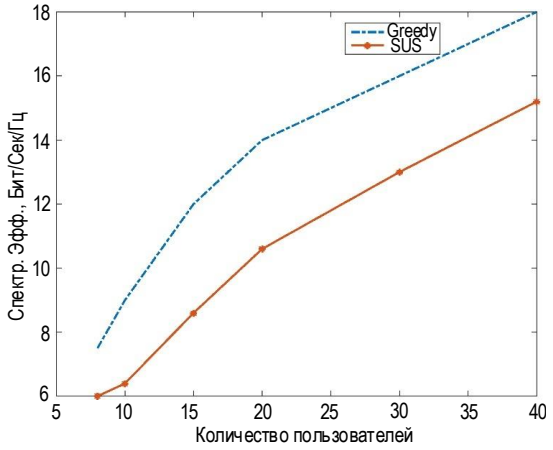


Рис. 4. Сравнение СЭ при прекодировании ZF и выборе абонентов SUS и Greedy

Fig. 4. Comparison of Sum Rate using ZF Precoding for User Selection SUS and Greedy

При увеличении количества абонентов *K* алгоритм Greedy ZF показывает заметный выигрыш в получаемой СЭ по сравнению с алгоритмом SUS.

Заключение

В статье представлены результаты моделирования алгоритма прекодирования ZF с выбором абонентов. Характер изменения СЭ определяется текущей пространственной корреляцией в канале, численно отображаемой величиной чисел обусловленности. Для указанного сценария распространения и параметров моделирования по полученным реализациям каналов пользователей прекодер ZF с выбором абонентов SUS показывает бóльшую СЭ по сравнению с прекодированием ZF и случайным выбором абонентов. СЭ увеличивается при увеличении количества выбираемых пользователей при количестве обслуживаемых пользователей бóльшем, чем количество антенн на стороне БС. При увеличении количества абонентов *K* алгоритм Greedy ZF показывает выигрыш в получаемой СЭ по сравнению с алгоритмом SUS.

Представленные результаты были получены в предположении равенства средних значений отношения сигнал/шум всех пользователей, что является следствием применения нормирования векторов канала при моделировании. При этом у выбранных пользователей реализуются равные значения средних спектральных эффективностей. При системном моделировании совместно с алгоритмом планирования следует учесть потери на распространение и различие средних отношений сигнал/шум пользователей.

Представленные результаты получены при условии статичности абонентов, для дальнейшего изучения свойств прекодирования и выбора абонентов необходимо включить в условия распространения перемещение абонентов, что приведет к появлению дополнительных ошибок при оценке канала, увеличению скорости изменения свойств канала и изменению получаемой спектральной эффективности. Также для дальнейшего изучения свойств метода выбора абонентов целесообразно провести моделирование более сложных алгоритмов прекодирования *BD* и регуляризованных алгоритмов.

Список источников

1. Bjornson E., Hoydis J., Sanguinetti L. Massive MIMO Networks: Spectral, Energy, and Hardware Efficiency // Foundations and Trends in Signal Processing. 2017. Vol. 11. Iss. 3–4. PP. 154–655. DOI:10.1561/2000000093

2. Castaneda E., Silva A., Gameiro A., Kountouris M. An Overview on Resource Allocation Techniques for Multi-User MIMO Systems // IEEE Communications Surveys and Tutorials. 2017. Vol. 19. Iss. 1. PP. 239–284. DOI:10.1109/COMST.2016.2618870

3. ETSI TS 38.211 V15.8.0 (2020-01). 5G; NR; Physical channels and modulation.

4. Bayesteh A., Khandani A.K. On the User Selection for MIMO Broadcast Channels // IEEE Transactions on Information Theory. 2008. Vol. 54. Iss. 3. PP. 1086–1107.

5. Yoo T., Goldsmith A. On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming // IEEE Journal on Selected Areas in Communications. 2006. Vol. 24. Iss. 3. PP. 528–541. DOI:10.1109/JSAC.2005.862421

6. Dimic G., Sidiropoulos N.D. On downlink beamforming with greedy userselection: Performance analysis and a simple new algorithm // IEEE Transactions on Signal Processing. 2005. Vol. 53. Iss. 10. PP. 3857–3868. DOI:10.1109/TSP.2005.855401

7. Kaltenberger F., Gespert D., Knopp R., Kountouris M. Performance of Multi-User MIMO Precoding with Limited Feedback over Measured Channels // Proceedings of the IEEE Global Telecommunications Conference (IEEE GLOBECOM 2008, New Orleans, USA, 30 November–04 December 2008). IEEE, 2008. DOI:10.1109/GLOCOM.2008.ECP.738
8. Cho Y.S., Kim J., Yang W.Y., Kang C.G. MIMO-OFDM Wireless Communications with MATLAB. John Wiley and Sons, 2010. 544 p.
9. ETSI TR 138.901 V16.1.0 (2020-11). 5G; Study on channel model for frequencies from 0,5 to 100 GHz.
10. Clerckx B., Kim G., Sung J. Correlated Fading in Broadcast MIMO Channels: Curse or Blessing? // Proceedings of the IEEE Global Telecommunications Conference (IEEE GLOBECOM 2008, New Orleans, USA, 30 November–04 December 2008). IEEE, 2008. DOI:10.1109/GLOCOM.2008.ECP.735
11. Jaeckel S., Raschkowski L., Börner K., Thiele L., Burkhardt F., Eberlein E. Quasi Deterministic Radio Channel Generator. User Manual and Documentation. QuaDRiGa. Document Revision: v2.2.0. Berlin: Fraunhofer Heinrich Hertz Institute, 2019.
12. Jaeckel S., Raschkowski L., Börner K., Thiele L. QuaDRiGa: A 3-D Multicell Channel Model with Time Evolution for Enabling Virtual Field Trials // IEEE Transactions on Antennas Propagation. 2014. Vol. 62. Iss. 6. PP. 3242–3256. DOI:10.1109/TAP.2014.2310220

References

1. Bjornson E., Hoydis J., Sanguinetti L. Massive MIMO Networks: Spectral, Energy, and Hardware Efficiency. *Foundations and Trends in Signal Processing*. 2017;11(3–4):154–655. DOI:10.1561/2000000093
2. Castaneda E., Silva A., Gameiro A., Kountouris M. An Overview on Resource Allocation Techniques for Multi-User MIMO Systems. *IEEE Communications Surveys and Tutorials*. 2017;19(1):239–284. DOI:10.1109/COMST.2016.2618870
3. ETSI TS 38.211 V15.8.0 (2020-01). *5G; NR; Physical channels and modulation*.
4. Yoo T., Goldsmith A. On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming. *IEEE Journal on Selected Areas in Communications*. 2006;24(3):528–541. DOI:10.1109/JSAC.2005.862421
5. Dimic G., Sidiropoulos N.D. On downlink beamforming with greedy userselection: Performance analysis and a simple new algorithm. *IEEE Transactions on Signal Processing*. 2005;53(10):3857–3868. DOI:10.1109/TSP.2005.855401
6. Dimic G., Sidiropoulos N.D. On downlink beamforming with greedy userselection: Performance analysis and a simple new algorithm. *IEEE Transactions on Signal Processing*. 2005;53(10):3857–3868. DOI:10.1109/TSP.2005.855401
7. Kaltenberger F., Gespert D., Knopp R., Kountouris M. Performance of Multi-User MIMO Precoding with Limited Feedback over Measured Channels. *Proceedings of the IEEE Global Telecommunications Conference, IEEE GLOBECOM 2008, 30 November–04 December 2008, New Orleans, USA*. IEEE; 2008. DOI:10.1109/GLOCOM.2008.ECP.738
8. Cho Y.S., Kim J., Yang W.Y., Kang C.G. *MIMO-OFDM Wireless Communications with MATLAB*. John Wiley and Sons; 2010. 544 p.
9. ETSI TR 138.901 V16.1.0 (2020-11). *5G; Study on channel model for frequencies from 0,5 to 100 GHz.*
10. Clerckx B., Kim G., Sung J. Correlated Fading in Broadcast MIMO Channels: Curse or Blessing? *Proceedings of the IEEE Global Telecommunications Conference, IEEE GLOBECOM 2008, 30 November–04 December 2008, New Orleans, USA*. IEEE; 2008. DOI:10.1109/GLOCOM.2008.ECP.735
11. Jaeckel S., Raschkowski L., Börner K., Thiele L., Burkhardt F., Eberlein E. Quasi Deterministic Radio Channel Generator. User Manual and Documentation. *QuaDRiGa*. Document Revision: v2.2.0. Berlin: Fraunhofer Heinrich Hertz Institute; 2019.
12. Jaeckel S., Raschkowski L., Börner K., Thiele L. QuaDRiGa: A 3-D Multicell Channel Model with Time Evolution for Enabling Virtual Field Trials. *IEEE Transactions on Antennas Propagation*. 2014;62(6):3242–3256. DOI:10.1109/TAP.2014.2310220

Статья поступила в редакцию 10.02.2023; одобрена после рецензирования 23.03.2023; принята к публикации 18.04.2023.
The article was submitted 10.02.2023; approved after reviewing 23.03.2023; accepted for publication 18.04.2023.

Информация об авторе:

КАЛАЧИКОВ Александр Александрович

кандидат технических наук, доцент кафедры радиотехнических систем Сибирского государственного университета телекоммуникаций и информатики

https://orcid.org/0000-0003-1235-6314

Научная статья
УДК 621.37
DOI:10.31854/1813-324X-2023-9-2-72-80

Модель самоорганизующейся сети радиосвязи, функционирующей в сложной сигнально-помеховой обстановке

Валерий Алексеевич Липатников, lipatnikovanl@mail.ru
Михаил Игоревич Петренко ✉, petrenko.m.i@mail.ru

Военная академия связи им. Маршала Советского Союза С. М. Буденного,
Санкт-Петербург, 194064, Российская Федерация

Аннотация: *Сети радиосвязи, в том числе применяющие адаптацию, предназначены для обмена информацией между отдельными корреспондентами и строятся, как правило, посредством радиотрасс, функционирующих в сложных условиях сигнально-помеховой обстановки. Необходимо учитывать степень влияния значений адаптивных параметров на показатели, описывающие соответствия требований к связи, энергетическую составляющую радиолинии, а также объем ресурсов радиолинии, затрачиваемые на ведение и восстановление связи. Получение оценок границ характеристик обслуживания трафика первичных и вторичных пользователей в самоорганизующейся радиосети, функционирующей в сложной сигнально-помеховой обстановке, является актуальным. Целью исследования является повышение достоверности результатов моделирования за счет получения граничных значений пропускной способности при передаче информации в самоорганизующейся сети радиосвязи. Проведено моделирование процессов, протекающих в сети радиосвязи, определены: граничные параметры задержки и загрузки; параметры выходного потока в сложных условиях сигнально-помеховой обстановки. Представлены выводы о достоинствах метода сетевого исчисления, по результатам серии проведенных вычислений. Получены аналитические оценки качества предоставления услуг в системе радиосвязи с использованием теории сетевого исчисления. Разработанная математическая модель позволяет исследовать показатели задержки, загрузки в самоорганизующейся сети радиосвязи при информационном обмене трафика различного вида в условии воздействия преднамеренных и непреднамеренных помех. Результаты аналитических расчетов, полученных при применении метода сетевого исчисления, могут быть использованы при формировании управляющих воздействий, а также решении задач повышения устойчивости радиолиний.*

Ключевые слова: *сетевое исчисление, модель, самоорганизующаяся сеть, поток, программно-определяемое радио, помехозащищенность, система массового обслуживания*

Ссылка для цитирования: Липатников В.А., Петренко М.И. Модель самоорганизующейся сети радиосвязи, функционирующей в сложной сигнально-помеховой обстановке // Труды учебных заведений связи. 2023. Т. 9. № 2. С. 72–80. DOI:10.31854/ 1813-324X-2023-9-2-72-80

Model of a Self-Organizing Radio Network, Operating in a Complex Signal and Interference Environment

Valery Липатников, lipatnikovanl@mail.ru
Mikhail Petrenko ✉, petrenko.m.i@mail.ru

Telecommunications Military Academy,
St. Petersburg, 194064, Russian Federation

Abstract: *Radio communication networks, including those that use adaptation, are designed for information exchange between individual correspondents and are usually built via radio routes, functioning ones in a complex signal and interference environment. It is necessary to take into account the degree of influence values adaptive parameters for indicators that describe accordance requirements communication requirements, the energy component of the radio link, as well as the amount of radio link resources spent on maintaining and restoring communication. Getting estimates of the boundaries of primary and secondary user traffic service characteristics in a self-organizing radio network, functioning in difficult signal-to-noise conditions it is relevant. The aim of the research is to increase the reliability of the simulation results by obtaining the boundary values of the throughput when transmitting information in a self-organizing radio communication network. Modeling of processes occurring in the radio communication network is carried out, and the following parameters are determined: boundary parameters of delay and loading; parameters of the output stream in a complex signal-interference environment. Conclusions about the advantages of the network calculus method based on the results of a series of calculations are presented. Analytical estimates of the quality of service provision in the radio communication system are obtained using the theory of network calculus. The developed mathematical model makes it possible to study the delay and load indicators in a self-organizing radio network during the information exchange of traffic of various types under the influence of intentional and unintentional interference. The results of analytical calculations obtained by applying the network calculus method can be used in the formation of control actions, as well as solving problems of increasing the stability of radio links.*

Keywords: *network calculus, model, self-organizing network, flow, software-defined radio, noise immunity, queuing system*

For citation: Lipatnikov V., Petrenko M. Model of a Self-Organizing Radio Network, Operating in a Complex Signal and Interference Environment. *Proc. of Telecom. Universities.* 2023;9(2):72–80. (in Russ.) DOI:10.31854/1813-324X-2023-9-2-72-80

Введение

Требования к своевременности и достоверности информационного обмена в радиолиниях (р/л), функционирующих в сложной сигнально-помеховой обстановке (ССПО), противоречия между требованиями к помехоустойчивости и пропускной способности (ПС) р/л способствуют поиску новых способов моделирования. Наиболее прогрессивным является способ организации р/л с применением элементов интеллектуализации, позволяющий разрешить/смягчить вышеуказанное противоречие [1]. Однако применение интеллектуальных способов обуславливает необходимость получения граничных оценок состояния масштабных р/л при низком уровне их наблюдаемости. Требуется разработать новые способы оценки, т. к. классический методический аппарат не позволяет обеспечить достаточную адекватность прогноза изменения характеристик масштабируемых р/л, функционирующих в ССПО. Известны методы повышения помехоустойчивости линий радиосвязи [2–4]: путем применения помехоустойчивого кодирования, увеличения базы сигнала и пространственной обработки сигналов. Известны способы повышения ПС линий радиосвязи [5]. Отметим, что в большинстве работ применяется методический аппарат, использующий классическую теорию массового обслуживания [6]. В работе [7] функционирование р/л рассматривается в условиях квазистационарности внешней среды. Кроме того, при разработке методик не учитываются свойства самоорганизации р/л. Одной из актуальных проблем, связанных с устойчивым функционированием самоорганизующейся сети радиосвязи в ССПО, является выполнение требований по их помехозащищенности. Решение может быть получено на основе синтеза алгоритмов адаптивного управления параметрами р/л, функционирующих в ССПО [8].

Объект исследования – технология когнитивного радио, позволяющая работать вторичным пользователям на частотных ресурсах первичных пользователей. Предмет исследования – граничные оценки характеристик качества обслуживания трафика в когнитивной радиосети в ССПО.

Цель: Повышение достоверности результатов моделирования за счет получения граничных значений ПС при передаче информации в самоорганизующейся сети радиосвязи.

Постановка задачи: Разработать модель самоорганизующейся сети для исследования зависимости характеристик входящего потока и качества обслуживания в условиях помех. Исследовать граничные задержки р/л, функционирующей в ССПО.

Исходные данные: Имеется самоорганизующаяся сеть радиосвязи с заданными параметрами, которая состоит из средств радиосвязи с изменяемыми рабочими характеристиками. Организация радиодоступа носимых терминалов осуществляется при помощи базовых станций различного частотно-территориального плана с временным разделением каналов. Настройки позволяют обеспечить передачу информации между корреспондентами в ССПО (рисунок 1).

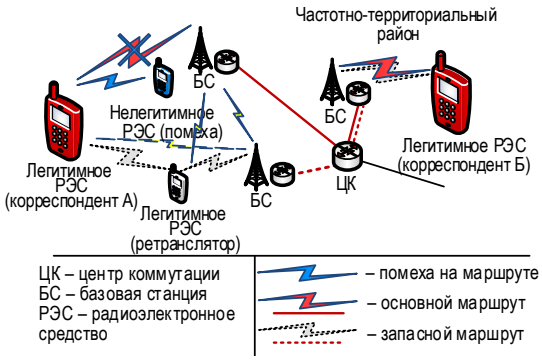


Рис. 1. Вариант построения самоорганизующейся сети радиосвязи

Fig. 1. Option for Building a Self-Organizing Radio Network

Введем следующие допущения относительно модели самоорганизующейся сети радиосвязи в ССПО:

1) известны: данные о характеристике радиоинтервала и режимах работы р/л, характеристики входного потока и дисциплины обслуживания FIFO (*аббр. от англ.* First-In First-Out, первый пришел – первый ушел), позиционирование места источников радиоизлучения;

2) имитация преднамеренных помех осуществляется на рабочей частоте РЭС (UE, *аббр. от* User Equipment), по одному сигналу (сигнально-кодовой конструкции);

3) работа алгоритмов управления р/л позволяет динамически изменять режимы работы РЭС;

4) р/л функционирует в соответствии с разработанными радиоданными.

С учетом сделанных допущений задача самоорганизации сводится к постоянному поддержанию требуемого качества канала в условиях ограниченного частотно-энергетического ресурса.

Решение задачи: Рассмотрено программно-определяемое РЭС как элемент р/л (рисунок 2).

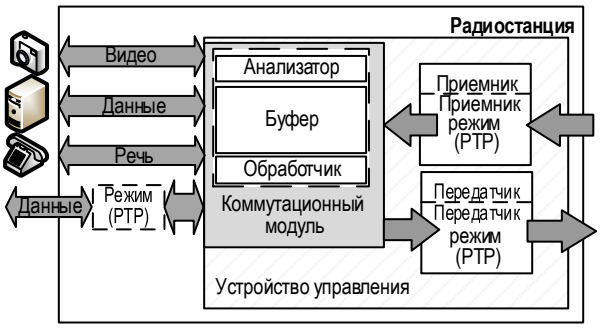


Рис. 2. Вариант функциональной схемы элемента радиолинии

Fig. 2. Option Functional Diagram of a Radio Link Element

Для разработки и исследования модели р/л выбран метод сетевого исчисления (*от англ.* Network Calculus) [9]. Метод позволяет представить процессы, протекающие в функциональной схеме радиостанции (см. рисунок 2), в виде модели (рисунок 3), состоящей из элементов классической системы массового обслуживания (СМО) с обратной связью. Функция $\alpha(t)$, возрастающая в широком смысле, является кривой поступления для функции входящего потока $A(t)$ тогда и только тогда, когда для $\forall \mathbf{0} \leq \boldsymbol{\tau} \leq \boldsymbol{t}$ справедливо неравенство: $A(t) - A(\tau) \geq \alpha(t - \tau)$. Имеется система обслуживания S и потоки трафика на входе и выходе этой системы $A(t)$ и $D(t)$, соответственно. Говорится, что система S реализует для потока $A(t)$ кривую обслуживания $\beta(t)$, если для любого момента времени $t \geq 0$ существует некоторое $t_0 \geq 0$, $t_0 \leq t$, такое, что $D(t) - A(t) \geq \beta(t - t_0)$.

Обслуживание заявок в СМО с обратной связью производятся по закономерностям стохастических (случайных) процессов, и, как следствие, традиционная теория массового обслуживания [10] не дает возможности аналитически получить характеристики качества обслуживания заявок (QoS, *аббр. от англ.* Quality of Service).

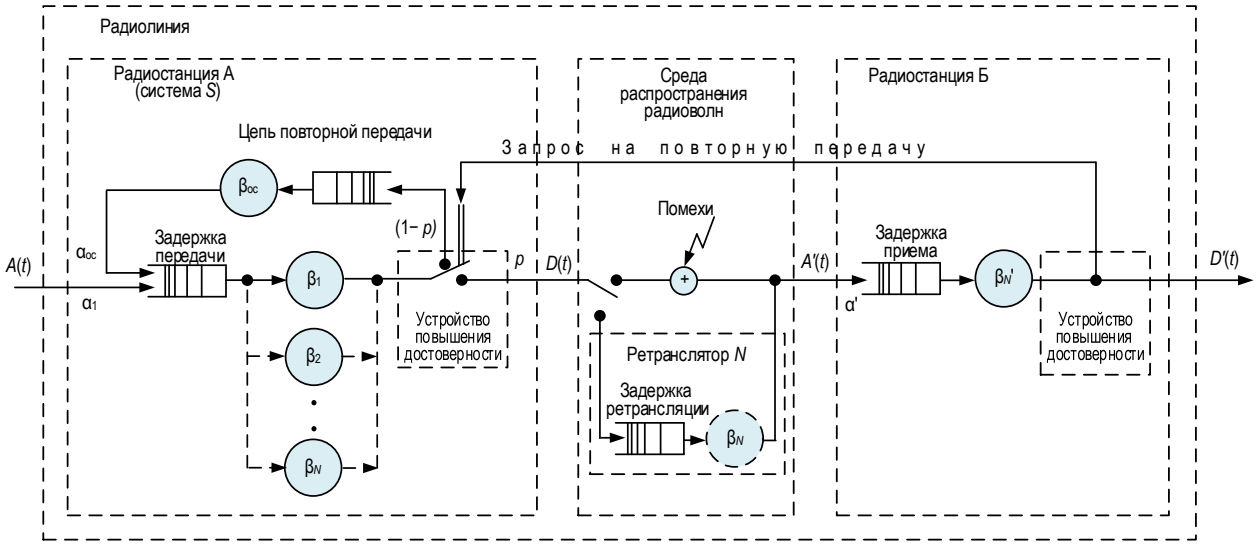


Рис. 3. Модель радиолинии в ССПО

Fig. 3. Radio Link Model in a Complex Signal and Interference Environment

Предполагается, что применение нового метода позволит получить ряд аналитических оценок QoS сложного потока. Ранее в классических моделях СМО с агрегацией трафика, проведение аналитических расчетов не проводилось из-за проблем, обусловленных наличием сложного потока заявок на входе системы *S*.

Алгоритм функционирования имитационной модели

Рассмотрим модель р/л в терминах сетевого исчисления [11], представленную на рисунке 3. При моделировании предполагается, что трафик передается в радиоэфир последовательно, по мере поступления в систему *S* с неограниченным буфером (без потерь). Поскольку потоки в телекоммуникационных сетях могут быть представлены в виде последовательности пакетов со сдвигом, то для их естественного описания лучше подходят дискретные модели. Моделирование проводится при значении параметров, удовлетворяющих достаточным условиям существования движения в системе [12]. Основные процессы, протекающие в модели приведены в виде отдельных блоков (рисунок 4).

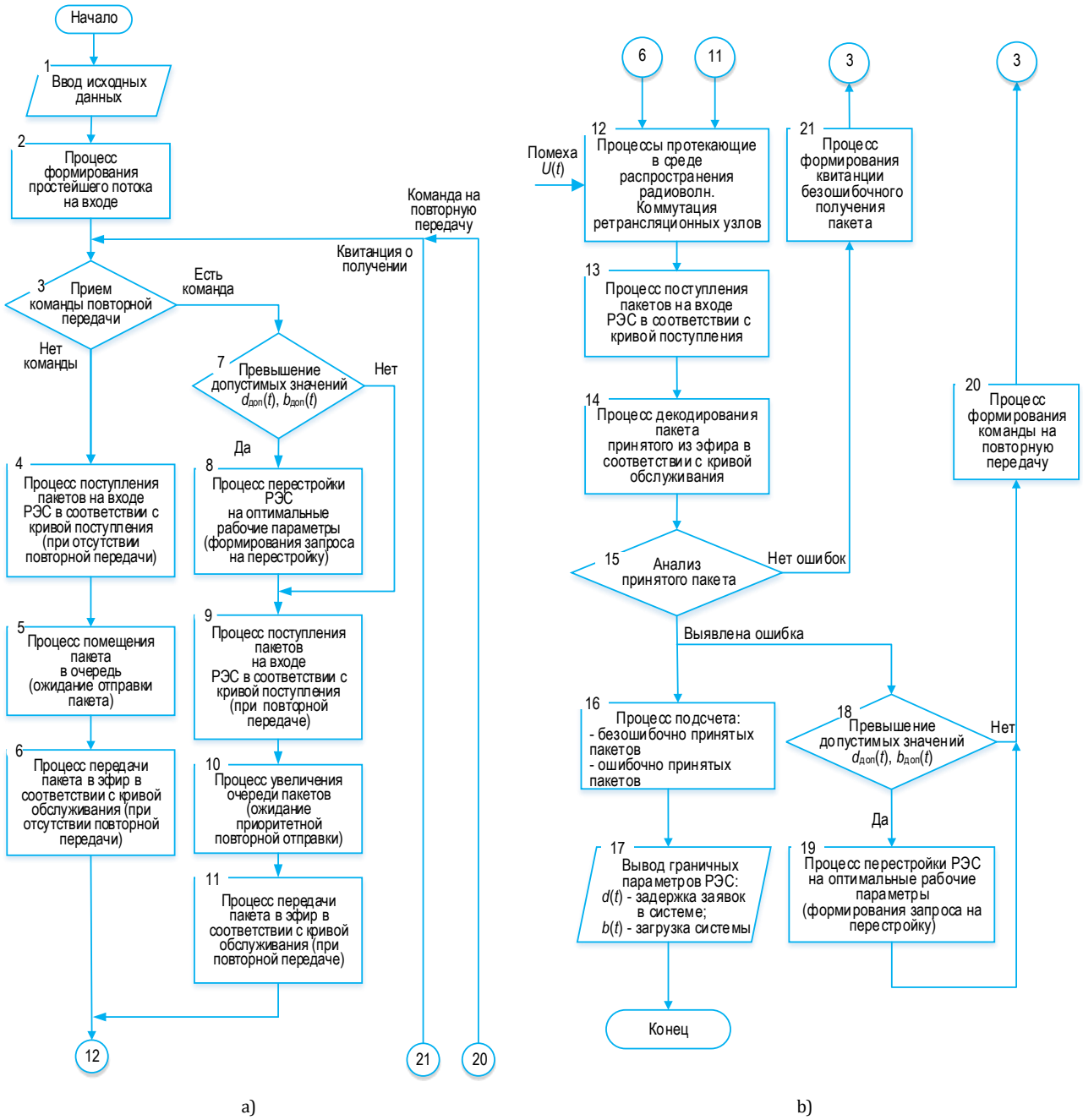


Рис. 4. Алгоритм функционирования модели: a) Радиостанция А; b) Радиостанция Б

Fig. 4. Operation Algorithm Models: a) Radio Station A; b) Radio Station B

1) Сначала задаются следующие входные параметры (блок 1): R – минимальная скорость обслуживания потока, бит/с; T – максимальная задержка потока; σ – предельная величина потока (берстность); ρ – устойчивая скорость потока; N – количество пакетов; $P_{ош}$ – вероятность ошибки.

2) Формируемый на входе системы простой поток заявок $\gamma_{\sigma,\rho}$ (блок 2) соответствует кривой поступления (блок 4) при условии отсутствия запросов на повторную передачу пакетов:

$$\alpha(t) = \gamma_{\sigma,\rho} = \begin{cases} \rho t + \sigma, t > 0 \\ 0, t \le 0 \end{cases}.$$

3) В модели предусмотрено устройство повышения достоверности передачи сообщений (блок 3), которое имеет возможность осуществлять повторную передачу пакета. В таком случае на входе будет формироваться сложный поток заявок:

$$\alpha(t) = \gamma_{\sigma,\rho} \wedge \gamma_{\sigma_{OC},\rho_{OC}} = \min\{\rho_N t + \sigma_N,\ \rho_{OC} t + \sigma_{OC}\}.$$

4) Очередной пакет помещается в бесконечный буфер, в соответствии с кривой поступления $\alpha(t)$. Процесс хранения пакета $\sum t_{задержки} = t_{дост} + t_{квит}$ (блок 5) продолжается до момента прихода на (блок 3) квитанции о безошибочном получении пакета, сформированной в блоке 21. Напротив, при приходе запроса на повторную передачу, устройство повышения достоверности (блоки 3, 15) осуществляет повторную передачу пакета, хранящегося в буфере, при этом процесс хранения продолжается до прихода следующей квитанции о правильном приеме от блока 21.

5) В соответствии с тактом работы модели осуществляется работа по передаче пакета в соответствии с кривой обслуживания:

– для случая безошибочной передачи (блок 6)

$$\beta(t) = \beta_{R,T} = \begin{cases} R(t-T), t > T \\ 0, \qquad\quad t \le T \end{cases},$$

– для других случаев (блок 11)

$$\beta_{oc}(t) = \{\beta_{R,T}\} \otimes \{\beta_{R_{OC},T_{OC}}\}.$$

6) В условиях ухудшения сигнально-помеховой обстановки [13] возрастает $P_{ош}$, а следом – задержка $d(t)$ и загрузка $b(t)$ в системе S. Блоки 7, 18 позволяют оценить величину задержки и загрузки с допустимыми: $d(t) \le d_{доп}(t)$ и $b(t) \le b_{доп}(t)$, а также произвести перестройку рабочих параметров РЭС и сформировать запрос на перестройку рабочих параметров РЭС корреспондента (блоки 8, 19).

Отличительной особенностью современных и перспективных средств связи является возможность программной перестройки рабочих параметров: например, выбор рабочей сигнально-кодовой конструкции (с соответствующей скоростью) $\Delta\beta_N$, способной обеспечить обработку и передачу радиокорреспонденту входящего потока посылок $\alpha(t)$ (рисунок 5).

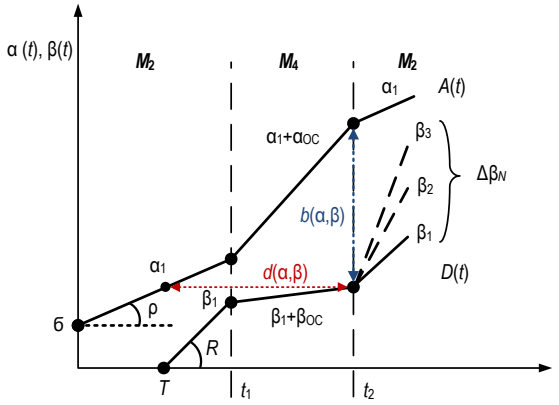


Рис. 5. Характеристики обслуживания потока при восстановлении связи M_4

Fig. 5. Characteristics of Thread Maintenance during Connection Recovery M_4

7) В процессе преодоления расстояния от одного корреспондента к другому в среде распространения радиоволн происходят процессы, оказывающие влияния на электромагнитный сигнал. В модели рассматривается процесс аддитивного влияния помехи $U(t)$ на информационный сигнал (блок 12). При формировании запроса на перестройку (блок 8) модель процесса функционирования р/л имеет возможность имитации работы ретрансляционных узлов (перестройки состава р/л).

8) Принятый сигнал поступает на бесконечный буфер приемного устройства со скоростью потока $\alpha'(t) = \gamma'_{\sigma,\rho}(t)$ (блок 13) и хранится там до момента полного декодирования, которое осуществляется со скоростью $\beta'(t) = \beta'_{R,T}(t)$ (блок 14).

9) В случае выявления ошибок в блоке 15, происходит процесс формирования запроса на повторную передачу (блок 20). В свою очередь происходит работа устройства повышения достоверности (блоки 3, 15), благодаря которому передающая радиостанция осуществляет повторную передачу пакета, хранящегося в буфере.

10) Процесс повторной передачи характеризуется появлением на входе системы сложного, агрегированного трафика (блок 9), состоящего из пакета, хранящегося в буфере $\alpha_1(i)$, и очередного пакета $\alpha_1(t)$:

$$\alpha(t) = \gamma_{\sigma,\rho} \wedge \gamma_{\sigma_{OC},\rho_{OC}} = \min\{\rho_N t + \sigma_N,\ \rho_{OC} t + \sigma_{OC}\}.$$

Данная последовательность действий позволяет моделировать процессы восстановления связи M_3, M_4. Стоит отметить, что пакет, хранящийся в очереди, имеет высший приоритет по отношению к пакетам, вновь поступающим в систему для передачи корреспонденту. Процесс хранения пакета $\sum t_{задержки} = \Delta t_{дост} + \Delta\ t_{пер} + t_{квит}$ (блок 10) продолжается до момента прихода на (блок 3) команды на повторную передачу пакета, сформированной в блоке 20.

11) Итерационный процесс передачи пакетов (блоки 6, 11) происходит, пока все N пакетов не будут приняты корреспондентом. Данная последовательность действий позволяет моделировать процессы восстановления связи M_1, M_2. Блок 16 осуществляет функцию счетчика подсчета безошибочно принятых пакетов. Пользуясь свойствами идемпотентной алгебры [14], производится вывод результатов задержки заявок в системе $d(t)$, загрузки системы $b(t)$ (блок 17). В таблице 1 приведены формулы расчета задержки, загрузки для системы S при различных ситуациях.

В целях моделирования процессов восстановления связи M_3 или M_4 был выбран ситуационный подход в описании поведенческой модели нелегитимного РЭС, которое осуществляет свою работу на лицензируемых частотах легитимных РЭС. Проведено имитационное моделирование функционирования р/л в ССПО [15], общие параметры которого представлены в таблице 2. Временная диаграмма – на рисунке 6.

В соответствии с алгоритмом функционирования модели р/л РЭС могут находиться в следующих устойчивых состояниях (рисунок 7):

– дежурный прием (M_0);

– вхождение в связь (M_1);

– ведение связи (M_2);

– восстановление связи без нарушения синхронизации (M_3);

– восстановление связи с нарушением синхронизации (M_4).

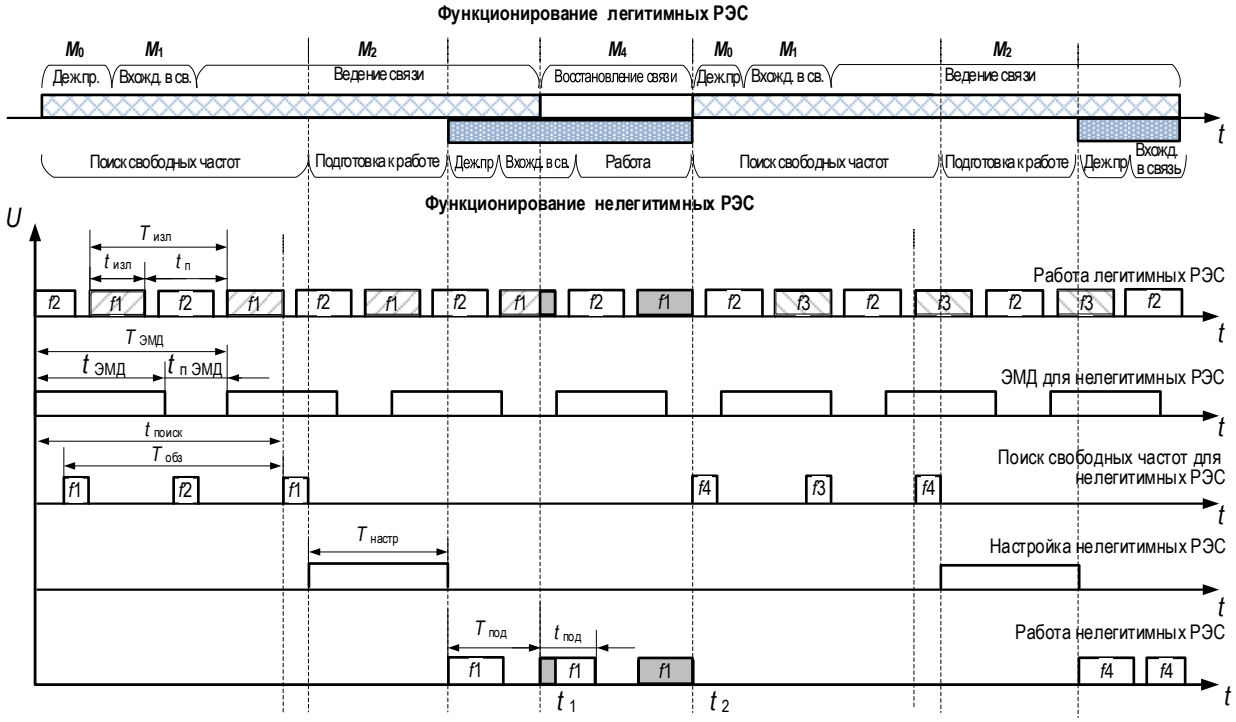


Рис. 6. Временная диаграмма функционирования РЭС

Fig. 6. Time Diagram of UE Operation

ТАБЛИЦА 1. Оцениваемые информационные показатели системы *S*

TABLE 1. Estimated Information Indicators of the System S

Показатели без обратной связи	Показатели с обратной связью
Задержка в системе	
$d(t)=\begin{cases}+\infty, & R<\rho\\ \dfrac{\sigma_N}{R_N+T_N}, & R\geq\rho\end{cases}$	$d(t)=\begin{cases}+\infty, & R<\rho\\ \sigma_N+\dfrac{\sigma_{OC}-\sigma_N}{\rho_N-\rho_{OC}}(\rho_N-R)^+ & \\ \dfrac{}{R_N\wedge R_{OC}}+(T_N,T_{OC}), & R\geq\rho\end{cases}$
Загрузка системы	
$b(t)=\begin{cases}+\infty, & R<\rho\\ \sigma_N+T_N\rho_N, & R\geq\rho\end{cases}$	$d(t)=\begin{cases}+\infty, & R<\rho\\ \sigma_N+\rho_{OC}\max\left(\dfrac{\sigma_{OC}-\sigma_N}{\rho_N-\rho_{OC}},(T_N,T_{OC})\right), & R\geq\rho\end{cases}$

ТАБЛИЦА 2. Общие параметры функционирования радиолинии

TABLE 2. General Parameters of Radio Link Operation

Обозначение	Описание
$f_1, f_2 \ldots f_N$	Лицензированные частоты для РЭС
$T_{изл}$	Период излучения легитимного f РЭС
$t_{изл}$	Время излучения легитимного f РЭС
$t_п$	Время перерыва между излучениями легитимного РЭС
$T_{ЭМД}$	Период электромагнитной доступности (ЭМД) легитимного РЭС
$t_{ЭМД}$	Время ЭМД работы легитимного РЭС
$t_{п\ ЭМД}$	Время перерыва в работе ЭМД легитимного РЭС
$T_{обз}$	Период обзора лицензионной частоты
$t_{поиск}$	Время поиска не занятой лицензионной частоты
$T_{настр}$	Время настройки не легитимного РЭС
$T_{под}$	Период подавления РЭС, работающего на лицензионной частоте
$t_{под}$	Время подавления РЭС, работающего на лицензионной частоте
M_0	Дежурный прием
M_1	Вхождение в связь
M_2	Ведение связи
M_3	Восстановление связи (без нарушения синхронизации)
M_4	Восстановление связи (с нарушением синхронизации)

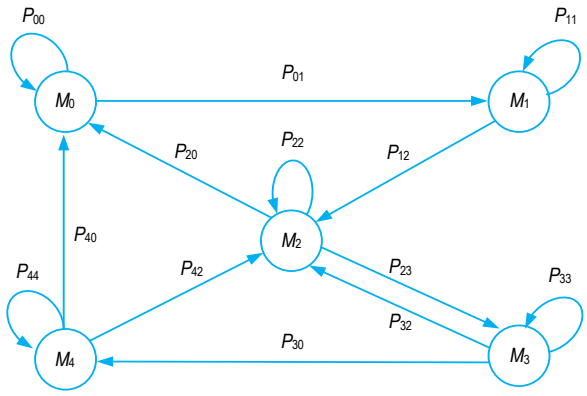


Рис. 7. Граф состояний радиолинии

Fig.7. Radio Link State Graph

Система переходит из одного состояния в другое с определенной вероятностью P, следовательно, можно воспользоваться аппаратом марковских случайных процессов, т. е. с помощью дифференциальных уравнений, в которых неизвестными являются P_0, P_1, P_2, P_3, P_4.

Контрольное решение

Пользуясь приведенными формулами, определим параметры выходного потока $\alpha'(t)$ и граничные параметры задержки $d(t)$, загрузки $b(t)$. Пусть на обслуживание поступает поток с постоянной скоростью 6,4 Мбит/с, для его обслуживания выбран режим работы, позволяющий вести обмен со скоростью от 9,6 до 19,2 Мбит/с (в зависимости от вида модуляции) (рисунок 8).

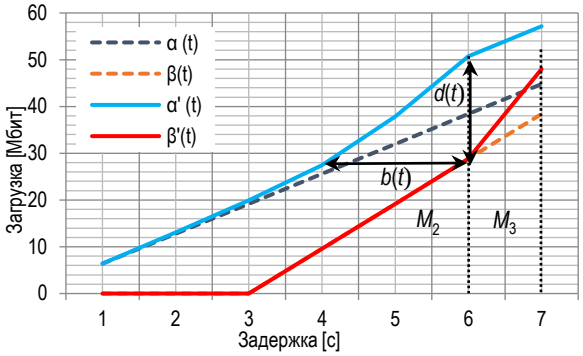


Рис. 8. Граничные характеристики загрузки буфера от задержки системы *S*

Fig. 8. Buffer Loading Boundary Characteristics as a Function of System Latency S

Для получения численных оценок рассматривается такт работы обслуживающего устройства, равным 1 секунде. Работа нелегитимного РЭС осуществляется за счет изменения величины $P_{ош}$ в пределах от 10^{-14} до 10^{-3}. В момент M_2 происходит включение нелегитимного РЭС (рисунок 6). При появлении помехи происходит изменение одного из исследуемых параметров $d_{max}(t) > d_{доп}(t)$ или $b_{max}(t) > b_{доп}(t)$, которое способствует нарушению устойчивой работы. Устройство управления фиксирует нарушение работы и приступает к реализации алгоритма поиска и настройки оптимальных рабочих параметров. На рисунке 8 представлены характеристики загрузки и обработки в системе S в моменты M_2, M_3.

Момент M_2 характеризуется увеличением загрузки буфера из-за появления запросов на повторную передачу, а M_3 – сменой основного маршрута на запасной. Запасной маршрут включает в свой состав ретрансляционный пункт (рисунок 1), который поддерживает работу на оптимальной сигнально-кодовой конструкции со скоростью 19,2 Мбит/с. Данные характеристики работы запасного маршрута позволяют быстрее опустошить переполненный буфер системы S.

На рисунке 9 представлены результаты аналитического моделирования с использованием выражения $d(t)$ (таблица 1). В момент завершения восстановления M_3 система S имеет задержку $d(t)$ не более 8 % от $d_{max}(t)$. Из анализа результатов на рисунке 9а следует, что метод сетевого исчисления позволяет исследовать зависимость задержки в системы S при динамической смене рабочих параметров р/л, функционирующей в ССПО. Стоит отметить, что многие современные системы связи используют показатель задержки для оценки QoS.

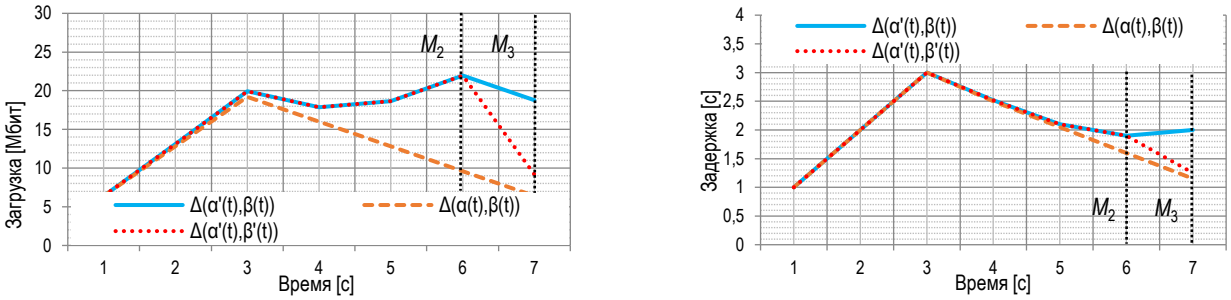


Рис. 9. Зависимость a) задержки и b) загрузки от времени моделирования радиолинии в ССПО

Fig. 9. The Delay (a) and Load (b) Dependence from the Simulation Time of Radio Line in a Complex Signal and Interference Environment

На рисунке 9b представлены результаты аналитического моделирования с использованием выражения $b(t)$ (таблица 1). В момент завершения восстановления M_3 система S имеет загрузку буфера $b(t)$ не более 18 % от $b_{max}(t)$. Из анализа результатов на рисунке 11 следует, что метод СИ позволяет исследовать зависимость загрузки буфера системы S при динамической смене рабочих параметров р/л, функционирующей в ССПО.

Новизна: Разработанная модель самоорганизующейся сети радиосвязи, функционирующая в ССПО, в отличие от известных, позволяет получить граничные значения информационной задержки и загрузки буфера при динамически изменяющихся параметрах р/л. *Практическая значимость:* Разработанная математическая модель позволяет исследовать показатели задержки, загрузки в самоорганизующейся сети радиосвязи при информационном обмене трафика различного вида в условии воздействия преднамеренных и непреднамеренных помех.

Заключение: Исследование процессов при помощи модели самоорганизующейся сети радиосвязи позволяет определить зависимость ПС р/л в условиях воздействия преднамеренных и непреднамеренных помех, а также – смены помехозащищенных режимов работы для обеспечения требуемой ПС.

Список источников

1. Липатников В.А., Парфиров В.А., Петренко М.И. Общая модель самоорганизующейся радиосвязи с мультиплексированием потоков // Международная научно-практическая конференция «Транспорт России: Проблемы и перспективы – 2022» (09–10 ноября 2022 г.). СПб.: Институт проблем транспорта им. Н.С. Соломенко РАН, 2022. Т. 1. С. 293–297.

2. Борисов В.И., Зинчук В.М., Лимарев А.Е. Помехозащищенность систем радиосвязи с расширением спектра сигналов модуляцией несущей псевдослучайной последовательностью. М.: Радио и связь, 2003. 640 с.

3. Рабин А.В. Помехоустойчивость систем цифровой связи с ортогональным кодированием и многопозиционной модуляцией. СПб.: ГУАП, 2019. 157 с.

4. Глушанков Е.И., Митянин С.А. Анализ совместной эффективности пространственно-временного кодирования и пространственной обработки сигналов в линиях радиосвязи // Заметки ученого. 2022. № 6. С. 187–192.

5. Дворников С.В., Манаенко С.С., Пшеничников А.В. Спектрально-эффективные сигналы с непрерывной фазой // Вестник Воронежского государственного технического университета. 2016. Т. 12. № 2. С. 87–93.

6. Липатников В.А., Сахаров Д.В., Парфиров В.А., Петренко М.И. Имитационная модель распределенного объекта радиоконтроля, отражающая динамику перемещений и смену режимов работы радиоэлектронных средств // Юбилейная XVIII Санкт-Петербургская международная конференция «Региональная информатика (РИ-2022)», Санкт-Петербург, Россия, 26–28 октября 2022 г. СПб: Санкт-Петербургское Общество информатики, вычислительной техники, систем связи и управления, 2022. С. 556–558.

7. Дворников С.В., Пшеничников А.В., Бурыкин Д.А. Структурно-функциональная модель сигнального созвездия с повышенной помехоустойчивостью // Информация и космос. 2015. № 2. С. 4–7.

8. Сорокин К.Н. Модель системы управления параметрами линии радиосвязи на основе нечеткой логики // Информация и космос. 2018. № 4. С. 39–43.

9. Фёдоров И.В., Росляков А.В. Анализ характеристик когнитивной радиосети с использованием сетевого исчисления // XXI Международная научно-техническая конференция «III научный форум телекоммуникации: теория и технологии, ТТТ-2019», Казань, Россия, 18–22 ноября 2019 г. Казань: Казанский государственный технический университет им. А.Н. Туполева, 2019. Т. 1. С. 338–339.

10. Белов А.В., Липатников В.А., Фёдоров И.В. Модель когнитивной радиосети на основе теории стохастического сетевого исчисления // X Международная научно-техническая и научно-методическая конференция «Актуальные проблемы инфотелекоммуникаций в науке и образовании (АПИНО 2021)», Санкт-Петербург, Россия, 24–25 февраля 2021 г. СПб.: СПбГУТ, 2021. Т. 1. С. 86–90.

11. Росляков А.В., Лысиков А.В., Витевский В.Д. Сетевое исчисление (Network Calculus). Часть 1. Теоретические основы // Инфокоммуникационные технологии. 2018. Т. 16. № 1. С. 19–33. DOI:10.18469/ikt.2018.16.1.02

12. Кудрявцева Е.Н., Росляков А.В. Базовые принципы и перспективы использования теории сетевого исчисления (Network Calculus) // Инфокоммуникационные технологии. 2013. Т. 11. № 3. С. 34–39.

13. Алекаев А.Е., Белов А.В., Фёдоров И.В. Способ многоступенчатой адаптации низкоэнергетической радиолинии коротковолнового диапазона с учетом прогнозирования сигнально-помеховой обстановки // Международная научно-практическая конференция «Транспорт России: Проблемы и перспективы – 2021» (Санкт-Петербург, Россия, 09–10 ноября 2021 г.). Т. 2. СПб.: Институт проблем транспорта им. Н.С. Соломенко РАН, 2021. С. 157–161.

14. Кривулин Н.К. Методы идемпотентной алгебры в задачах моделирования и анализа сложных систем. СПб.: СПбГУТ, 2009. 256 с.

15. Пшеничников А.В. Оценка статистических параметров рабочих частот функциональных моделей радиолиний в конфликтной ситуации // Информация и космос. 2018. № 1. С. 46–50.

References

1. Lipatnikov V.A., Parfirov V.A., Petrenko M.I. General model of self-organizing radio communication with stream multiplexing. *Proceedings of the International Scientific and Practical Conference on Transport of Russia: Problems and Prospects – 2022, 09–10 November 2022, St. Petersburg, Russia, vol.1*. St. Petersburg: IPT RAN Publ.; 2022. p.293–297. (in Russ.)

2. Borisov V.I., Zinchuk V.M., Limarev A.E. Noise Immunity of Radio Communication Systems with Signal Spectrum Expansion by Carrier Pseudorandom Sequence Modulation. Moscow: Radio and Communications Publ.; 2003. 640 p. (in Russ.)

3. Rabin A.V. Noise immunity of digital communication systems with orthogonal coding and multi-position modulation. St. Petersburg: Saint Petersburg State University of Aerospace Instrumentation Publ.; 2019. 157 p. (in Russ.)

4. Glushankov E.I., Mityanin S.A. Analysis of the joint efficiency of spatial-time coding and spatial processing of signals in radio communication lines*. Zametki uchenogo.* 2022;6:187–192. (in Russ.)

5. Dvornikov S.V., Dvornikov S. S., Manaenko S.S., Pshenichnikov A.V. Spectral-efficient signals with the continuous phase. *Bulletin of Voronezh State Technical University*. 2016;12(2):87–93. (in Russ.)

6. Lipatnikov V.A., Sakharov D.V., Parfirov V.A., Petrenko M.I. Simulation model of a distributed radio monitoring object, reflecting the dynamics of movements and changing modes of operation of radio-electronic means. *Proceedings of the Jubilee XVIII St. Petersburg International Conference on Regional Informatics (RI-2022), St. Petersburg, Russia, 26–28 October 2022*. St. Petersburg: Sankt-Peterburgskoe Obshchestvo informatiki vychislitelnoi tekhniki sistem sviazi i upravleniia Publ.; 2022. p.556–558. (in Russ.)

7. Dvornikov S.V., Pshenichnikov A.V., Burykin D.A. Structural and functional model of a signal constellation with increased noise immunity. *Information and Space*. 2015;2:4–7. (in Russ.)

8. Sorokin K. N. Model of a radio link parameter management system based on fuzzy logic. *Information and Space*. 2018;4: 39–43. (in Russ.)

9. Fedorov I.V., Roslyakov A.V. Analysis of the characteristics of a cognitive radio network using network calculus. *Proceedings of the XXIst International Scientific and Technical Conference "III Scientific Forum of Telecommunications: Theory and Technology", TTT-2019, 18–22 November 2019, Kazan, Russia, vol.1*. Kazan: KAI Publ.; 2019. p.338–339. (in Russ.)

10. Belov A., Lipatnikov V., Fedorov I. Model of a cognitive radio network based on the theory of stochastic network calculation. *Proceedings of the Xth International Scientific and Technical and Scientific-Methodical Conference on Actual Problems of Infotelec Communications in Science and Education, 24–25 February 2021, St. Petersburg, Russia, vol.1*. St. Petersburg: The Bonch-Bruevich Saint Petersburg State University of Telecommunications Publ.; 2021. p.86–90.p. (in Russ.)

11. Roslyakov A.V., Lysikov A.V., Vitevsky V.D. Network Calculus. Part 1. Theoretical Foundations. *Infokommunikacionnye tehnologii*. 2018;16(1):19–33. (in Russ.) DOI:10.18469/ikt.2018.16.1.02

12. Kudryavtseva E.N., Roslyakov A.V. Basic principles and prospects of network calculus application. *Infokommunikacionnye tehnologii. 2013;11(3):34–39. (in Russ.)*

13. Alekaev A.E., Belov A.V., Fedorov I.V. Method for multi-stage adaptation of low-energy radio line of short-wave range taking into account forecasting signal-interference situation. *Proceedings of the International Scientific and Practical Conference on Transport of Russia: Problems and Prospects – 2021, 09–10 November 2021, St. Petersburg, Russia, vol.2.* St. Petersburg: Solomenko Institute of Transport Problems of the Russian academy of sciences Publ.; 2021. p.157–161. (in Russ.)

14. *Krivulin N.K. Meth*ods of idempotent algebra in problems of modeling and analysis of complex systems. *St. Petersburg: The Bonch-Bruevich Saint Petersburg State University of Telecommunications Publ*.; 2009. 256 p. (in Russ.)

15. Pshenichnikov A.V. Estimation of statistical parameters operating frequency of functional models of radio links in a conflict situation. *Information and Space*. 2018;1:46–50. (in Russ.)

Статья поступила в редакцию 30.01.2023; одобрена после рецензирования 27.02.2023; принята к публикации 15.03.2023.
The article was submitted 30.01.2023; approved after reviewing 27.02.2023; accepted for publication 15.03.2023.

Информация об авторах:

ЛИПАТНИКОВ Валерий Алексеевич — доктор технических наук, профессор, старший научный сотрудник научно-исследовательского центра Военной академия связи им. Маршала Советского Союза С.М. Буденного
https://orcid.org/0000-0002-3736-4743

ПЕТРЕНКО Михаил Игоревич — адъюнкт научно-исследовательского центра Военной академия связи им. Маршала Советского Союза С.М. Буденного
https://orcid.org/0000-0002-5402-402X

Научная статья
УДК 004.27
DOI:10.31854/1813-324X-2023-9-2-81-93

Интегральное решение проблемы размещения контроллеров и балансировки нагрузки

Аммар Салех Али Мутханна, muthanna.asa@sut.ru

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

Аннотация: *Наиболее эффективным методом построения ядра сетей связи пятого и последующих поколений в настоящее время представляется использование мультиконтроллерных программно-конфигурируемых сетей SDN. К настоящему времени существует целый ряд алгоритмов для размещения контроллеров в мультиконтроллерных сетях, основанных на метаэвристических методах вследствие сложности решаемых задач и алгоритмов балансировки нагрузки, позволяющих обеспечить наилучшее использование их ресурсов. Однако интегрального решения проблемы размещения контроллеров и балансировки нагрузки пока найдено не было. Именно решению такой проблемы и посвящена настоящая статья. С целью достижения поставленной цели в работе предложено совместно использовать кластеризацию сети и метаэвристический хаотический алгоритм «роя сальп», хорошо зарекомендовавший себя в предыдущих исследованиях по проблемам построения мультиконтроллерных сетей. С учетом интегрального решения проблемы размещения контроллеров на базе кластеризации мультиконтроллерной сети и балансировки нагрузки алгоритм «роя сальп» в статье модифицирован. Анализ эффективности предложенного решения проведен путем сравнения результатов моделирования как с широко известными метаэвристическими алгоритмами «роя частиц» (PSO) и «серого волка» (GWO), так и с предыдущей версией хаотического алгоритма «роя сальп» (CSSA).*

Ключевые слова: *сети связи пятого и последующих поколений, ядро сети, мультиконтроллер, балансировка нагрузки, алгоритм «роя частиц» (PSO), алгоритм хаотического «роя сальп» (CSSA), алгоритм «серого волка» (GWO), кластеризация*

Источник финансирования: Работа выполнена в рамках прикладных научных исследований СПбГУТ, регистрационный номер 1022040500653-0 от 16.02.2023 в ЕГИСУ НИОКТР.

Ссылка для цитирования: Мутханна А.С.А. Интегральное решение проблемы размещения контроллеров и балансировки нагрузки // Труды учебных заведений связи. 2023. Т. 9. № 2. С. 81–93. DOI:10.31854/1813-324X-2023-9-2-81-93

Controller Location and Load Balancing Integrated Solution

Ammar Muthanna, muthanna.asa@sut.ru

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

Abstract: *The usage of multi-controller SDNs is currently the most efficient approach for constructing the core of communication networks of the fifth and following generations. Due to the complexity of the problems being tackled, there are currently a number of load balancing algorithms and algorithms for arranging controllers in multi-controller networks that are based on meta-heuristic methods. These algorithms allow for the optimum possible utilisation of controller resources in such networks. However, a comprehensive solution to the load balancing and controller placement issues has yet to be discovered. The answer to such an issue is the focus of this article. The report suggests using network clustering in conjunction with the meta-heuristic chaotic salp swarm technique, which has shown to be effective in prior research on the challenges of creating multi-controller networks, to accomplish this goal. The salp swarm algorithm in the paper is adjusted to take into account the integral solution to the problem of*

deploying controllers based on clustering of a multi-controller network and load balancing. By contrasting the simulation results with those from the well-known meta-heuristic particle swarm algorithms optimization and the grey wolf GWO, as well as the previous version of the chaotic salp swarm algorithm CSSA, the effectiveness of the proposed solution was evaluated.

Keywords: *5G networks, core network, multi-controller, load balancing, chaotic salp swarm algorithm, particle swarm optimization algorithms, grey wolf optimization algorithms, clustering*

Funding: The research was supported by Applied Scientific Research under the SPbSUT state assignmen No. 1022040500653-0, 2023.

For citation: Muthanna A. Controller Location and Load Balancing Integrated Solution. *Proc. of Telecom. Universities.* 2023;9(2):81–93. (in Russ.) DOI:10.31854/1813-324X-2023-9-2-81-93

1. Введение

Развитие сетей и систем связи в направлении создания сетей связи пятого и последующих поколений ставит все новые и новые задачи перед исследовательскими центрами во всем мире. Появление концепций Интернета Вещей и Тактильного Интернета [1] привело к созданию сверхплотных сетей и сетей связи с ультрамалыми задержками, принципиально изменивших представления о трафике, поступающем на сети, и о требованиях к качеству обслуживания в таких сетях. Это привело к необходимости пересмотра представлений о построении ядра сети. При этом наиболее подходящей новой технологией для построения сетей связи пятого и последующих поколений были признаны программно-конфигурируемые сети SDN (*аббр. от англ.* Software-Defined Network) [2, 3].

Научной проблемой, исследуемой в статье, является разработка алгоритма, обеспечивающего интегральное решение для оптимального размещения контроллеров в мультиконтроллерных сетях, основанных на метаэвристических методах вследствие сложности решаемых задач, и балансировки нагрузки, позволяющей обеспечить наилучшее использование ресурсов контроллеров таких сетях.

Основной вклад данной статьи заключается в:

– разработке алгоритма иерархической кластеризации мультиконтроллерной сети для решения проблемы интеграции размещения контроллеров в мультиконтроллерных сетях и балансировки нагрузки;

– разработке модифицированного алгоритма хаотического «роя сальп» (CSSA, аббр. от англ. Chaotic Salp Swarm Algorithm – хаотический алгоритм «роя сальп») для использования в иерархических кластерных сетях clus-CSSA.

Статья организована следующим образом: в разделе (2) приводится аналитическая информация о проблеме размещения контроллеров и балансировки нагрузки в мультиконтроллерных сетях SDN, в разделе (3) представлена разработанная кластерная архитектура для балансировки нагрузки в мультиконтроллерных сетях, в разделе (4) – сетевая модель исследуемой мультиконтроллерной сети SDN; в разделе (5) представлена решаемая оптимизационная задача, в разделе (6) – оценка характеристик разработанных решений с использованием кластеризации и метаэвристического модифицированного хаотического «роя сальп» в сравнении с другими известными алгоритмами.

2. История вопроса и соответствующие работы

Размещение контроллеров в SDN является одной из критических проблем и привлекает большое внимание в литературе [4]. Проблема размещения контроллеров в мультиконтроллерных сетях впервые была исследована в [5], а затем и [6].

В [7] авторы разработали метод балансировки нагрузки для динамического добавления контроллеров в заданную сеть, при этом коммутаторы могут мигрировать между контроллерами в зависимости от нагрузки на последние. Была представлена идея динамического назначения между коммутатором и контроллером и предложен эффективный алгоритм для балансировки нагрузки и миграции коммутаторов. Однако в исследовании не было учтено существенное влияние первоначального размещения контроллеров на их загрузку.

В [8] авторы модифицировали задачу *k*-центра для повышения эффективности распределения ресурсов между клиентами. Клиентам выделяются ограниченные ресурсы с учетом ограничений на пропускную способность соответствующих объектов. Формулируется оптимизационная задача и вычисляется решение с помощью линейного и смешанного целочисленного программирования, когда в заданном радиусе доступно необходимое количество объектов с достаточной пропускной способностью.

В [9] авторы предложили метаэвристические решения для размещения контроллеров с использованием алгоритмов оптимизации «роя частиц» (PSO, *аббр. от англ.* Particle Swarm Optimization) и светлячков и сравнили результаты со случайным размещением контроллеров. Результаты моделирования показывали, что оба алгоритма работают лучше с точки зрения задержки и более быстрой сходимости. В приведенном анализе результатов

для индийской сети TATA число контроллеров составляет 20. Однако оптимальность наблюдаемого числа контроллеров не была определена.

В работе [10] авторы представили решения на основе алгоритма игры с ненулевой суммой для оптимального размещения нескольких контроллеров. В игре с ненулевой суммой каждый контроллер имеет механизм оптимизации, который вычисляет функцию вознаграждения и сравнивает свое собственное значение вознаграждения для экономии затрат и улучшения качества обслуживания (QoS, *аббр. от англ.* Quality of Service) путем оптимизации расположения контроллеров. Аналогично, в [11] авторы представили игровую модель для изучения размещения нескольких контроллеров. Эта модель учитывает несколько метрик, которые включают задержку связи между контроллерами и коммутаторами, накладные расходы на связь между контроллерами и нагрузку на них. На основе этих метрик в статье сформулирована оптимизационная задача с двумя противоречивыми целями: минимизация задержки связи и накладных расходов на связь.

В [12] авторы предложили использовать теорию очередей для размещения нескольких контроллеров. Был использован CSSA для решения задачи оптимизации. Авторы провели сравнение разработанного алгоритма с другими метаэвристическими алгоритмами. Результаты моделирования показали, что предложенный алгоритм превосходит другие метаэвристические алгоритмы и алгоритм на основе теории игр по своим характеристикам.

В [13] авторы сформулировали задачу оптимизации для размещения нескольких контроллеров, учитывая сценарий отказа одного контроллера в SDN. В работе был предложен метаэвристический алгоритм имитационного отжига для достижения глобального оптимального решения. Однако предварительное определение числа контроллеров привело к неоптимальному их размещению в динамически изменяющейся сети реального времени.

В [14] авторы предложили алгоритм разбиения сети, основанный на лувенской эвристической методике определения сообществ. Для идентификации сообществ они вычислили гаверсово расстояние между ребрами сети и присвоили ребрам вес, обратно пропорциональный расстоянию. Для определения расположения контроллеров были реализованы отдельные алгоритмы, которые находят компромисс между средней задержкой и нагрузкой на контроллер. Однако сложность алгоритмов значительно возрастает с размером сети.

Алгоритмы, основанные на *k*-means, эвристике, Парето и многоцелевой оптимизации для размещения нескольких контроллеров, не решаются за требуемое время. Для решения проблемы делаются определенные допущения и приближения. В целях разработки простых в вычислительном отношении решений авторы работы [15] предложили разделение сети и оптимизированную кластеризацию *k*-means для разделения сети на *k* подсетей, когда речь идет о задержке. Техника разбиения сети применяется для упрощения проблемы размещения контроллеров. По сравнению с приведенными выше схемами подход на основе кластеризации *k*-средних снижает вычислительную сложность. Однако для алгоритма требуются предопределенные входные параметры, что делает его неприменимым в сценариях сетей, работающих в реальном времени.

В [16] авторы применили иерархическую кластеризацию *k*-means для разделения сети и расширили оптимизированную кластеризацию *k*-means, предложенную в [15], учитывая как задержку, так и балансировку нагрузки. Было использовано расстояние кратчайшего пути между узлами вместо евклидова расстояния, используемого в оптимизированном *k*-means. Алгоритм итеративно объединяет *k*′ начальных центров и в итоге получает *k* центров. Предварительно заданные значения *k* и *k*′ в иерархической кластеризации делают алгоритм неприменимым для глобальных вычислительных сетей (WAN, *аббр. от англ.* Wide Area Network) на базе SDN.

В [17] авторы разработали систему BalanceFlow, которая представляет собой типичное решение кластеризации контроллеров, основанное на иерархическом развертывании. Основным преимуществом этого метода является гибкая настройка запросов потока, обрабатываемых каждым контроллером, без введения дополнительных задержек распространения. Он следует функции мультиконтроллеров в OpenFlow 1.2. Все контроллеры в BalanceFlow поддерживают свою собственную информацию о нагрузке и периодически публикуют ее друг другу через систему межконтроллерной связи. При изменении состояния трафика один из контроллеров BalanceFlow выбирается в качестве суперконтроллера, который разделяет трафик и перераспределяет различные настройки потока между соответствующими контроллерами.

Новизна предлагаемой работы заключается в рассмотрении балансировки нагрузки с размещением контроллеров для мультиконтроллерных сетей SDN. Развертывание новой схемы кластеризации с новым разработанным метаэвристическим алгоритмом, т. е. CSSA, является основной новизной нашей работы, которая решает обе проблемы мультиконтроллерных сетей SDN. Насколько известно, это первая работа, в которой рассматриваются решения обеих проблем: размещение контроллеров и балансировка нагрузки.

3. Иерархическая кластеризация для мультиконтроллерных сетей SDN

Предложенный алгоритм иерархической кластеризации выполняет ее в динамическом режиме, основываясь на сетевом трафике и запросах потоков. Плоскость управления делится на кластеры SDN-контроллеров с головным SDN-контроллером для каждого. На рисунке 1 показана структура кластерной сети SDN. Сеть SDN состоит из централизованного альфа-контроллера (C_α) и кластеров обычных SDN-контроллеров с равным их числом в каждом кластере. В каждом кластере есть головной узел, который представляет бета-контроллер (C_β) и отвечает за настройку кластера и балансировку нагрузки между контроллерами, входящими в него. Основная цель разработанного алгоритма кластеризации – сбалансировать нагрузку между контроллерами плоскости управления и избежать отказов контроллеров. Это снижает стоимость сети и повышает ее общую доступность и надежность.

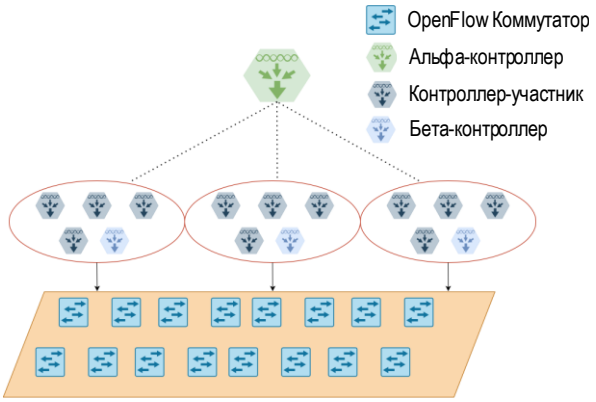


Рис. 1. Общая структура кластерной мультиконтроллерной сети SDN

Fig. 1. A Clustered Multi-Controller SDN Network's General Structure

Процесс кластеризации делится на фазы 1) настройки контроллера, 2) соединения коммутаторов и 3) установившегося состояния. На этапе создания кластера альфа-контроллер формирует SDN-кластеры, выбирая бета-контроллеры и контроллеры-участники каждого кластера. Контроллер с минимальной ожидаемой нагрузкой выбирается в качестве бета-контроллера и берет на себя роль руководителя членов кластера. Все кластеры формируются однородно, т. е. количество SDN-контроллеров в каждом кластере одинаково. После формирования кластеров фаза настройки контроллеров заканчивается, и начинается фаза настройки соединений коммутаторов. На этапе установки соединений коммутаторов вызывается разработанный CSSA, представленный в разделе 5.

Каждому кластеру назначается группа OpenFlow-коммутаторов, и, кроме того, CSSA динамически распределяет коммутаторы между каждым контроллером-участником таким образом, чтобы достичь оптимальной эффективности задержки и стоимости, как CAPEX, так и OPEX. На этом этапе каждый SDN-контроллер принимает роль, т. е. либо члена кластера, либо бета, и назначает оптимальные соединения с OpenFlow-коммутаторами, определяемые CSSA. Затем начинается фаза установившегося состояния. Каждый контроллер-член кластера управляет подключенными OpenFlow-коммутаторами и берет на себя роль таблицы потоков таких коммутаторов. Каждый бета-контроллер управляет назначенными OpenFlow-коммутаторами как контроллеры-члены и, кроме того, берет на себя роль руководителя своего кластера.

Бета-контроллер отслеживает трафик между узлами-участниками своего кластера и отправляет отчеты альфа-контроллеру SDN. Когда бета-контроллер обнаруживает дисбаланс нагрузки среди членов кластера, например, некоторые контроллеры перегружены более, чем на 90 % от максимальной нагрузки, или другие контроллеры недогружены – менее, чем на 30 % от максимальной нагрузки, он выполняет перекластеризацию, называемую межкластеризацией. Процесс межкластеризации направлен на балансировку нагрузки среди контроллеров-участников путем перемещения роли головного узла на контроллер-участник с минимальной нагрузкой, который становится новым бета-контроллером, и вызова CSSA для переподключения SDN-контроллеров кластера к OpenFlow-коммутаторам. Процесс межкластерного объединения состоит из трех фаз, как и общая кластеризация, однако бета-контроллер выполняет межкластерное объединение, в отличие от общей кластеризации, которая выполняется альфа-контроллером.

Работа сети продолжается до тех пор, пока нагрузка между кластерами не станет несбалансированной. Альфа-контроллер получает отчеты от бета-контроллеров и отслеживает трафик между различными кластерами; когда он обнаруживает дисбаланс нагрузки между ними, он выполняет новый раунд общей кластеризации, формируя новые кластеры с помощью ранее введенных фаз. Эта схема поддерживает балансировку нагрузки между распределенными SDN-контроллерами, что позволяет достичь оптимального использования сети.

4. Математическое моделирование кластерной мультиконтроллерной сети SDN

Набор развернутых SDN-контроллеров в плоскости управления – это вектор C, и он определяется следующим образом:

$$C = \{C_1, C_2, C_3, \dots \dots \dots \dots, C_N\},\ \ C_\alpha \in C,\ C_\beta \subset C, \qquad (1)$$

где N – общее количество развернутых SDN-контроллеров.

Общее количество развернутых кластеров равно M, и оно отличается от раунда к раунду. Набор бета-контроллеров определяется следующим образом:

$$C_\beta = \left\{C_{\beta_1}, C_{\beta_2}, C_{\beta_3}, \dots\dots\dots, C_{\beta_M}\right\}, \quad C_\beta \subset C \quad \forall\, C_{\beta_i} \in C. \tag{2}$$

Каждый сформированный кластер имеет набор развернутых контроллеров-членов, длина которого равна L. Набор контроллеров-членов для каждого кластера определяется по выражению:

$$C_{mi} = \{C_{mi,1}, C_{mi,2}, C_{mi,3}, \dots\dots\dots, C_{mi,L}\}, \quad C_{mi} \subset C \quad \forall\, C_{mi,j} \in C. \tag{3}$$

В плоскости данных сети SDN развернуто k OpenFlow-коммутаторов, распределенных между кластерами контроллеров. Каждый коммутатор имеет соединение с SDN-контроллером, который распределяется с помощью разработанного алгоритма размещения контроллеров, представленного в следующем разделе. Набор развернутых OpenFlow-коммутаторов определяется выражением:

$$S = \{S_1, S_2, S_3, \dots\dots\dots, S_K\}. \tag{4}$$

Каждый кластер SDN-контроллеров имеет набор подключенных коммутаторов, которые можно определить как:

$$S_{mi} = \{S_{mi,1}, S_{mi,2}, S_{mi,3}, \dots\dots\dots, S_{mi,R}\}, \quad S_{mi} \subset S \quad \forall\, S_{mi,j} \in S. \tag{5}$$

Соединения между коммутаторами и SDN-контроллерами указываются в матрице коммутации, где строки указывают на SDN-контроллер, а столбцы вводятся для OpenFlow-коммутаторов, при этом матрица T представляет общее количество подключенных коммутаторов на каждый SDN-контроллер.

Пример матрицы коммутации представлен в виде:

$$[0 \cdots 1 \vdots \ddots \vdots 1 \cdots 0], T = [2 \vdots 3]. \tag{6}$$

Одним из способов проверки производительности контроллера является оценка временного отклика контроллера, на который в основном влияет задержка в очереди. Контроллеры могут быть смоделированы с помощью многосерверной модели очередей $M/M/s$, где предполагается, что каждый контроллер имеет s ядер [18]. Передаваемые пакеты поступают на контроллер с определенной скоростью, соответствующей процессу Пуассона, образуя единую очередь на контроллере.

Среднее время ответа T_i контроллера C_i представляет собой сумму времени ожидания в очереди и времени обработки и может быть рассчитано по формуле Erlang C как функция скорости поступления λ_i и скорости обслуживания μ:

$$T_i(\lambda) = \frac{C\left(s, \frac{\lambda_i}{\mu}\right)}{s\mu_i - \lambda_i} + \frac{1}{\mu}, \tag{7}$$

где $C(s, \lambda/\mu)$ – вероятность того, что все серверы системы используются, и любой прибывающий пакет будет поставлен в очередь, и может быть рассчитан как в уравнении:

$$C\left(s, \frac{\lambda}{\mu}\right) = \frac{\left(\frac{(s\rho)^c}{s!}\right)\left(\frac{1}{1-\rho}\right)}{\sum_{k=0}^{s-1}\frac{(s\rho)^k}{k!} + \left(\frac{(s\rho)^c}{s!}\right)\left(\frac{1}{1-\rho}\right)} = \frac{1}{1 + \left(\frac{1}{1-\rho}\right)\left(\frac{s!}{(s\rho)^c}\right)\sum_{k=0}^{s-1}\frac{(s\rho)^k}{k!}}, \tag{8}$$

$$\rho = \frac{\lambda_i}{s\mu}, \tag{9}$$

где ρ – коэффициент использования сервера, который является показателем стабильности системы.

Система имеет стабильное распределение только в том случае, если утилизация серверов ρ меньше единицы. Когда поступивших заявок в очереди больше, чем серверов контроллера, переход все равно будет только с $s\mu$ и не более, а контроллер будет находиться в максимальной пропускной способности.

Скорость прибытия контроллера может быть рассчитана как сумма средних скоростей прибытия коммутаторов, подключенных к контроллеру:

$$\lambda_i = \sum_{k_i} \lambda_s \tag{10}$$

Средняя нагрузка на контроллер C_i может быть рассчитана как среднее количество запросов, поставленных в очередь и обработанных. Используя (7), средняя нагрузка на контроллер Li может быть рассчитана по формуле (11):

$$L_i(\lambda) = s\rho + \frac{\rho}{1-\rho} C\left(s, \frac{\lambda_i}{\mu}\right). \tag{11}$$

5. Проблема размещения контроллеров с точки зрения задержки и эффективности затрат

5.1. Формулировка проблемы

В связи с динамическим изменением нагрузки на сеть проблема размещения контроллера должна решаться динамически – чтобы достигались определенные показатели сети. В этом разделе эта проблема сформулирована с точки зрения задержки, использования и стоимости сети. Задача размещения контроллеров при этом направлена на получение динамически оптимального числа SDN-контроллеров, которое обеспечивает требуемую задержку между распределенными OpenFlow-коммутаторами и SDN-контроллерами, а также минимальную стоимость сети. Это оптимальное количество динамически изменяется в зависимости от нагрузки на сеть. Более того, задача направлена на определение оптимальных соединений между коммутаторами и SDN-контроллерами таким образом, чтобы

достичь требуемой латентности и стоимости. Модифицируем ранее разработанную нами задачу размещения контроллеров с учетом задержки и стоимости, представленную в [12], для кластерной сети SDN.

Задача размещения контроллеров определяется таким образом, чтобы распределить новые контроллеры или сократить существующие в соответствии с динамическим изменением сетевого трафика. Проблема оптимизации формулируется для получения оптимального количества SDN-контроллеров и оптимального количества контроллеров-участников на кластер таким образом, чтобы достичь эффективности задержки, стоимости, использования и балансировки нагрузки. Задача оптимизации представляет собой функцию минимизации, целью которой является уменьшение общего числа SDN-контроллеров в сети N_T, общего числа SDN-контроллеров на кластер N_C, общей стоимости SDN-контроллеров, включающей как CAPEX, так и OPEX, и средней задержки между контроллером и подключенными коммутаторами, включающей распространение, постановку в очередь и обработку.

Задача формулируется следующим образом:

$$\text{Min } f(N_T, N_C, C, D). \tag{12}$$

Ограничения:

$$T_i^j \leq T_{\text{thr}}, \quad \forall i \epsilon \gamma \ \wedge \ j \epsilon \beta, \tag{13}$$

$$U_{lb} \leq U_i^j \leq U_{ub}, \quad \forall i \epsilon \gamma \ \wedge \ j \epsilon \beta, \tag{14}$$

$$U_{C-lb} \leq U_C^j \leq U_{C-ub}, \quad \forall j \epsilon \beta, \tag{15}$$

где f – нелинейная функция общего количества развернутых SDN-контроллеров N_T, общего количества SDN-контроллеров на кластер N_C; C – общая стоимость SDN-контроллеров, включающая CAPEX и OPEX; D – средняя задержка между контроллером и подключенными коммутаторами, включающая распространение, постановку в очередь и обработку.

Задача представляет собой многоцелевую оптимизацию с множеством ограничений, которая может быть решена соответствующим метаэвристическим алгоритмом с тремя ограничениями.

Первое указывает, что среднее время ответа T_i^j контроллера C_i, принадлежащего кластеру j, должно быть меньше порогового значения T_{thr}, которое является предопределенным значением; это имеет место для всех SDN-контроллеров в наборе доступных контроллеров [γ] для набора развернутых кластеров [β]. T_{thr} предопределено таким образом, чтобы удовлетворить требуемое QoS.

Второе вводится для поддержания уровня использования каждого SDN-контроллера. Показатель использования U_i^j контроллера C_i в кластере j должен находиться в пределах, с одной стороны, нижней границы использования SDN-контроллера U_{lb} (т. е. минимального значения использования SDN-контроллера, ниже которого он должен быть отключен для достижения экономической эффективности), а с другой – верхней (т. е. максимального значения использования SDN-контроллера, выше которого контроллер может быть перегружен). Верхний и нижний пределы U_{ub} и U_{lb} предопределены таким образом, чтобы достичь требуемого QoS. Этот показатель использования SDN-контроллера применяется для сопоставления с показателями использования мощности, хранения и обработки данных.

Третье учитывает поддержание общего индекса использования кластера j между максимальным (U_{C-ub}) и минимальным пределами (U_{C-lb}). Это ограничение введено для поддержания баланса нагрузки между кластерами и для предотвращения выхода из строя раздела сети.

5.2. Использование системы

В этом подразделе представлена фитнес-функция для сформулированной задачи. Это функция использования, которая применяется для сравнения различных решений и указания на лучшее решение путем отображения переменных или событий на реальные числа:

$$U: V \rightarrow R. \tag{16}$$

Использование времени – это первая функция полезности, которая применяется для отображения временной реакции SDN-контроллеров. Функции потерь хорошо подходят для данного случая, поэтому для моделирования использования времени можно применить любую их форму. Здесь рассматривается квадратичная функция, так как она математически проста благодаря своей симметрии.

Временная полезность SDN-контроллера C_i, принадлежащего кластеру j, равна $U^j_{T\text{-}Ci}$ и определяется следующим образом:

$$U_{T-Ci}^j = \{\alpha + \delta\left(T_{\text{thr}} - T_i^j(\lambda)\right)^2, \quad T_i^j(\lambda) \leq T_{\text{thr}} \ \ 0, T_i^j(\lambda) > T_{\text{thr}} \quad \forall i \epsilon \gamma \ \wedge \ j \epsilon \beta, \tag{17}$$

где α и δ – константы, значения которых не влияют на решение. Первой константе α может быть присвоено определенное значение, которое представляет собой минимальное ненулевое значение времени использования $U^j_{T\text{-}Ci}$, возникающее, когда время реакции равно пороговому значению.

Эти константы могут быть определены следующим образом:

$$\alpha = U_{T-\text{thr}} \quad \forall U_T \in [0,1], \tag{18}$$

$$\beta = \frac{(1 - U_{T-\text{thr}})}{T_{\text{thr}}^2} \quad \forall U_T \in [0,1]. \tag{19}$$

Для T_{thr} = 70 % время использования может быть пересчитано следующим образом:

$$U_{T-Ci}^{j} = \{0{,}7 + \frac{0{,}3}{T_{\text{thr}}^{2}}\left(T_{\text{thr}} - T_{i}^{j}(\lambda)\right)^{2}, \quad (20)$$

$$T_{i}^{j}(\lambda) \le T_{\text{thr}}\ \ 0, T_{i}^{j}(\lambda) > T_{\text{thr}} \quad \forall\, i\epsilon\,\gamma\ \wedge\ j\epsilon\,\beta.$$

Функция использования времени кластера j – U^{j}_{T} может быть рассчитана как нормализованное среднее значение использования времени каждого члена кластера (21).

Второй функцией полезности, которую следует рассмотреть, является функция полезности затрат, которая отображает стоимость используемых контроллеров. Под стоимостью в основном понимаются оба термина: CAPEX и OPEX, связанные с развернутыми контроллерами. Квадратичная функция потерь также представляет собой подходящую функцию для выражения утилизации затрат. Стоимость использования каждого контроллера в наборе доступных контроллеров может быть определена по выражению (22), где $U^{j}_{C\text{-}Ci}$ – функция использования затрат контроллера C_i в кластере j, а ρ – константа, которая не влияет на решение, независимо от того, какое значение ей присвоено. Правильное значение ρ может быть определено следующим образом, для $U^{j}_{C\text{-}Ci}$ между 0 и 1 (23). Функция использования затрат в кластере j – U^{j}_{C} и может быть рассчитана как нормализованное среднее значение использования затрат каждого члена кластера (24).

$$U_{T}^{j} = \frac{\sum_{N_C} U_{T-Ci}^{j}}{N_C} = \frac{\left(\sum_{N_C}\{0{,}7 + \frac{0{,}3}{T_{\text{thr}}^{2}}\left(T_{\text{thr}} - T_{i}^{j}(\lambda)\right)^{2},\ T_{i}^{j}(\lambda) \le T_{\text{thr}} 0\ ,\ T_{i}^{j}(\lambda) > T_{\text{thr}} \quad \forall\, i\epsilon\,\gamma\ \wedge\ j\epsilon\,\beta\right)}{N_C}, \quad (21)$$

$$U_{C-Ci}^{j} = \{\rho\left(U_{ub} - U_{i}^{j}\right)^{2} \quad \forall\ U_{i}^{j} \in [U_{lb}\ , U_{ub}] \quad 0 \quad \forall\ U_{i}^{j} \notin [U_{lb}\ , U_{ub}] \quad \forall\, i\epsilon\,\gamma\ \wedge\ j\epsilon\,\beta, \quad (22)$$

$$\rho = \frac{1}{(U_{ub} - U_{lb})^{2}} \quad \forall\, U_{C} \in [0{,}1] \quad (23)$$

$$U_{C}^{j} = \frac{\sum_{N_C} U_{\text{C}-Ci}^{j}}{N_C} = \frac{\left(\sum_{N_C}\{\rho(U_{ub} - U_{i}^{j})^{2} \quad \forall\ U_{i}^{j} \in [U_{lb}\ , U_{ub}] \quad 0 \quad \forall\ U_{i}^{j} \notin [U_{lb}\ , U_{ub}] \quad \forall\, i\epsilon\,\gamma\ \wedge\ j\epsilon\,\beta\right)}{N_C} \quad (24)$$

Общая полезность каждого развернутого SDN-контроллера может быть рассчитана как взвешенная сумма полезности времени и затрат, и таким образом можно оценить общую полезность каждого кластера.

Общая полезность контроллера C_i в j-м кластере равна U^{j}_{Ci} и рассчитывается следующим образом:

$$U_{Ci}^{j} = \delta_{C} U_{C-Ci}^{j} + \delta_{T} U_{T-Ci}^{j}, \quad (25)$$

где δ_C и δ_T – весовые коэффициенты затрат и времени, соответственно.

Функция общего использования системы U_{cont} представляет собой среднее значение использования каждого контроллера и рассчитывается следующим образом:

$$U_{cont} = \frac{\sum_{\beta}\sum_{N_C} U_{Ci}^{j}}{|\gamma|}. \quad (26)$$

Более того, средняя полезность затрат $U_{C\text{-}cont}$ и средняя полезность времени $U_{T\text{-}cont}$ всех доступных контроллеров могут быть рассчитаны следующим образом:

$$U_{C-cont} = \frac{\sum_{\beta}\sum_{N_C} U_{\text{C}-Ci}^{j}}{|\gamma|}, \quad (27)$$

$$U_{T-cont} = \frac{\sum_{\beta}\sum_{N_C} U_{T-Ci}^{j}}{|\gamma|}. \quad (28)$$

Общая полезность каждого сформированного кластера может быть рассчитана таким же образом, как и для отдельных контроллеров. Общая полезность кластера j равна U^{j} и может быть рассчитана как взвешенная сумма полезностей времени и затрат кластера j:

$$U^{j} = \delta_{C} U_{C}^{j} + \delta_{T} U_{T}^{j}. \quad (29)$$

Функция среднего использования каждого сформированного кластера U_{clus} и рассчитывается следующим образом:

$$U_{Clus} = \frac{\sum_{\beta} U^{j}}{|\beta|}. \quad (30)$$

Более того, средняя полезность затрат $U_{C\text{-}clus}$ и средняя полезность времени $U_{T\text{-}clus}$ всех сформированных кластеров вычисляются по выражениям:

$$U_{C-clus} = \frac{\sum_{\beta} U_{C}^{j}}{|\beta|}, \quad (31)$$

$$U_{T-clus} = \frac{\sum_{\beta} U_{T}^{j}}{|\beta|}. \quad (32)$$

5.3. Хаотический алгоритм «роя сальп» для кластерной сети SDN

В этом подразделе представляется процедура оптимизации на основе CSSA для решения задачи размещения контроллеров в кластерной сети SDN с в постановке, сформулированной в подразделе 5.1.

Разработанный алгоритм является модифицированной версией CSSA, представленного в [12].

Алгоритм CSSA – это метаэвристический популяционный алгоритм, имитирующий поведение сальпы в океанах. Это недавний тип оптимизаторов PSO, который моделирует поведение живых роев в реальной жизни. Рой сальп состоит из сальпы-лидера и сальп-последователей, которые движутся по цепочке вслед за лидером к позиции пищи, которая является наилучшей для них. Позиция сальпы моделируется как d-мерное пространство поиска, где d представляет собой количество переменных в определенной задаче, точно так же, как и другие алгоритмы на основе роя.

Вектор позиции n сальп в пространстве поиска имеет вид $X^j = [x_1^j, x_2^j, x_3^j, \ldots\ldots, x_d^j]$, $j = 1, 2, \ldots, n$, и лидер обновляет свою позицию, используя следующее уравнение:

$$X_i^1 = \{F_i + C_1\big((ub_i - ul_i)C_2 + lb_i\big), \quad C_3 \geq 0 \; F_i - C_1\big((ub_i - ul_i)C_2 + lb_i\big), \quad C_3 < 0, \tag{33}$$

где X_i^1 обозначает положение сальпы-лидера в измерении i^{th}; F_i – положение пищи в измерении i^{th}; ub_i и lb_i – верхняя и нижняя границы в измерении i^{th}; C_1, C_2, и C_3 – коэффициенты модели (представляют собой случайные числа, которые используются для решения определенных задач).

Первый коэффициент C_1 вводится для обеспечения баланса между разведкой и эксплуатацией, представляет собой наиболее важный параметр в алгоритме и определяется следующим образом:

$$C_1 = 2\mathrm{e}^{-\left(\frac{4t}{\mathrm{T_{max}}}\right)^2}, \tag{34}$$

где t – текущая итерация: $\mathrm{T_{max}}$ – максимальное число итераций.

C_2 и C_3 – случайные числа, которые равномерно генерируются со значениями от 0 до 1. Сальпы-последователи обновляют свои позиции на основе закона движения Ньютона, используя следующее уравнение:

$$X_i^k = \frac{1}{2}\left(X_i^k + X_i^{k-1}\right) \quad 2 \leq k < n, \tag{35}$$

где X_i^k – положение k^{th} последователей сальпа в i^{th} измерении; n – общее число частиц сальп.

Чтобы избежать спада в локальных оптимумах и низкой скорости сходимости, мы вводим хаотические карты в рассматриваемый оптимизатор «роя сальп». Хаотические карты вводятся для обновления оптимизатора вместо случайных чисел.

Используем логистическую карту для настройки значения второго коэффициента C_2 следующим образом:

$$C_2^t = \omega(t), \tag{36}$$

$$\omega(t+1) = a\omega(t)[1 - \omega(t)], \quad a = 4, \tag{37}$$

где $\omega(t)$ – значение логистической карты на итерации t^{th}, с начальным условием 0,7, т. е. $\omega(0) = 0{,}7$.

Для *NP*-трудной задачи, смоделированной в подразделе 5.1, мы стремимся определить оптимальное количество SDN-контроллеров и кластеров, а также оптимальное соединение для каждого коммутатора в наборе коммутаторов S. Путем итераций несколько роев сальп параллельно ищут оптимальные решения, которые представляют собой оптимальное количество контроллеров, и кластеров, а также оптимальное распределение контроллеров между коммутаторами. Следовательно, для получения оптимальных решений вводятся три вложенных алгоритма, основанных на ранее представленном хаотическом сальпе.

Алгоритм 1 представляет псевдокод CSSA, разработанного для задачи размещения контроллеров, поставленной в подразделе 5.1, где каждая сальпа представляет контроллер в сети. Выходом этого алгоритма является оптимальное количество SDN-контроллеров. Алгоритм 2 представляет псевдокод CSSA, разработанного для оптимального соединения SDN OpenFlow-коммутаторов на основе оптимального количества контроллеров, рассчитанного алгоритмом 1. Каждая сальпа в алгоритме 2 представляет все доступные соединения для всех коммутаторов с их выделенными контроллерами, и это M-мерный вектор, каждое измерение которого представляет коммутатор. Выходом второго алгоритма является наилучшее распределение контроллеров между коммутаторами. Алгоритм 3 представляет собой псевдокод CSSA, разработанного для оптимальной кластеризации. Выходом алгоритма 3 является оптимальное число кластеров.

Алгоритм 1. CSSA для поиска оптимального числа контроллеров

Initialize u_b, l_b, $T_{\max}$, d, n
Initialize positions of salps x_i (i = 1, 2, 3, …, n)
While ($t \leq t_{\max}$)
    Calculate the fitness function of each salp position using Eq. (15)
    F = The best salp position
    Update the value of C_1 using (23)
    Get the value of chaotic map (Logistic) $w(t)$
    For (i = 1: $i \leq n$) do
      if (i == 1)
        Update the position of the leading salp using (22, 23, 25, 26)
      else
        Update the position of the follower salps using (24)
      end if
    end for
    Adjust salps based on the upper and lower bounds
    Calculate the best connections of switches for the best salp (call Algorithm 2)
    Update the best salp based on the results of Algorithm 2
    Calculate the best number of clusters (call Algorithm 3)
    $t \leftarrow t+1$
Return F

Алгоритм 2. CSSA для оптимального соединения переключателей с контроллерами

Initialize *ub*, *lb*, T_{max}, *d*, *n*
Initialize positions of salps x_i (i = 1, 2, 3, ..., n)
While ($t \le t_{max}$)
  Calculate the fitness function of each salp position using (15)
  F = The best salp position
  Update the value of *C*1 using (23)
  Get the value of chaotic map (Logistic) $w(t)$
  For (i = 1: $i \le n$) do
    if (i == 1)
      Update the position of leading salp using (22, 23, 25, 26)
    else
      Update the position of follower salp using (24)
    end if
  end for
  Adjust salps based on the upper and lower bounds
  $t \leftarrow t$+1
Return F

Алгоритм 3. CSSA для оптимальной кластеризации

Initialize *ub*, *lb*, T_{max}, *d*, *n*
Initialize positions of salps x_i (i = 1, 2, 3, ..., n)
While ($t \le t_{max}$)
  Calculate the fitness function of each salp position using (19)
  F = The best salp position
  Update the value of *C*1 using (23)
  Get the value of chaotic map (Logistic) $w(t)$
  For (i = 1: $i \le n$) do
    if (i == 1)
      Update the position of leading salp using (22, 23, 25, 26)
    else
      Update the position of follower salp using (24)
    end if
  end for
  Adjust salps based on the upper and lower bounds
  $t \leftarrow t$+1
Return F

6. Оценка эффективности

6.1. Настройка моделирования

Разработанная оптимизированная схема кластеризации смоделирована в среде Matlab с использованием процессора Intel Core i7 и оперативной памяти 8 ГБ. Для оценки производительности рассмотрим реальные топологии крупномасштабных сетей WAN, чтобы лучше проиллюстрировать эффективность предложенной схемы. Для моделирования выбрано 15 приближенных реальных топологий из набора данных Internet Topology Zoo. Эти топологии рассматриваются как крупномасштабные сети с массивными устройствами пересылки. Эти выбранные топологии представлены в таблице 1 с их характеристиками. Каждая точка присутствия в топологии сети рассматривается как сетевой коммутатор, т. е. устройство пересылки. Предполагается, что поток является случайной величиной, подчиняющейся распределению Пуассона. В таблице 2 представлены рассматриваемые параметры моделирования. В качестве SDN-контроллера используется NOX-контроллер.

ТАБЛИЦА 1. Топологии сети, рассмотренные для моделирования

TABLE 1. Network Topologies Considered for Simulation

Ссылка. Номер	Топология	Количество
1	IBM	18
2	Oxford	20
3	FCCN	23
4	AGIS	25
5	Viatel	83
6	GEANT	27
7	TATA	169
8	GTS CE	187
9	PalmettoNet	49
10	OTEGlobe	78
11	DFN	47
12	GARR	36
13	ULAKNET	76
14	RNP	28
15	Carnet	43

ТАБЛИЦА 2. Параметры моделирования

TABLE 2. Simulation Parameters

Параметр	Описание	Количество
λ_S	Средняя скорость запроса коммутатора	[1500, 3000]
μ	Скорость обслуживания контроллера (запрос/сек)	30000
T_{thr}	Порог задержки (мс)	2
U_{ub}	Верхняя граница индекса использования контроллера	0.9
U_{lb}	Нижняя граница индекса использования контроллера	0.7
$U_{C\text{-}ub}$	Верхняя граница индекса использования кластера	0.9
$U_{C\text{-}lb}$	Нижняя граница индекса использования кластера	0.7
T_{max} (1)	Максимальное количество итераций для алгоритма 1	50
T_{max} (2)	Максимальное количество итераций для алгоритма 2	30
T_{max} (3)	Максимальное количество итераций для алгоритма 3	30
δ_C	Весовой коэффициент затрат	18
δ_T	Весовой коэффициент времени	25

6.2. Результаты моделирования

В интересах оценки производительности рассматриваются три сценария моделирования для исследования влияния на оптимальные топологические решения (число SDN-контроллеров) следующих параметров: 1) порога среднего времени отклика SDN-контроллера T_{thr} (Сценарий I), 2) верхнего индекса использования U_{ub} (Сценарий II), верхней границы индекса использования кластера $U_{C\text{-}ub}$ (Сценарий III). Для Сценариев I и III – U_{ub} установлена на 0.9; для Сценариев II и III – T_{thr} установлен на 2 мс; для Сценариев I и II – $U_{C\text{-}ub}$ установлена на 0.9. Разработанная оптимизированная схема кластеризации реализуется для каждой из пятнадцати топологий глобальной сети для четырех случаев (наборов параметров), представленных в таблице 3.

ТАБЛИЦА 3. Четыре набора параметров для трех сценариев моделирования

TABLE 3. Four Sets of Parameters for Three Simulation Scenarios

Ссылка. Номер	Сценарий I	Сценарий II	Сценарий III
Случай(1)	$T_{thr\text{-}1}$ = 1 мс	U_{ub1} = 0.8	$U_{C\text{-}ub1}$ = 0.8
Случай (2)	$T_{thr\text{-}2}$ = 2 мс	U_{ub2} = 0.90	$U_{C\text{-}ub2}$ = 0.0
Случай (3)	$T_{thr\text{-}3}$ = 3 мс	U_{ub3} = 0.2	$U_{C\text{-}ub3}$ = 0.2
Случай (4)	$T_{thr\text{-}4}$ = 4 мс	U_{ub4} = 0.4	$U_{C\text{-}ub4}$ = 0.4

Результаты показывают, что оптимальное количество активных SDN-контроллеров уменьшается по мере увеличения порогового времени отклика и индекса максимального использования каждого SDN-контроллера, а значит, уменьшается и оптимальное количество кластеров. С увеличением порогового времени отклика и индекса максимальной загрузки SDN-контроллер обрабатывает большее количество устройств пересылки. Это происходит за счет задержки, которая должна поддерживать требуемое QoS системы и нагрузки на контроллер, которая увеличивает вероятность сбоя и, кроме того, потребляет больше энергетических ресурсов.

Разработанная оптимизированная схема кластеризации реализована для случайной сети с плоскостью данных из 100 коммутаторов, чтобы оценить производительность CSSA для сетей большого масштаба. Основной целью данного сценария моделирования является оценка влияния вариации параметров сети на производительность оптимизированной схемы кластеризации. Производительность системы измеряется в широком диапазоне порогового времени обработки SDN-контроллера T_{thr} верхнего индекса использования SDN-контроллера U_{ub} и верхней границы индекса использования кластера $U_{C\text{-}ub}$.

Изменение этих параметров является критическим и оказывает значительное влияние на оптимальные топологические решения; особенно для крупномасштабных сетей с плотным развертыванием; более того – большое влияние на QoS сети.

Далее для моделирования рассматриваются три сценария с десятью случаями. В первом сценарии оценивается влияние изменения порогового времени обработки SDN-контроллера T_{thr}, во втором сценарии – верхнего индекса использования SDN-контроллера U_{ub}, а в третьем сценарий – верхней границы индекса использования кластера $U_{C\text{-}ub}$.

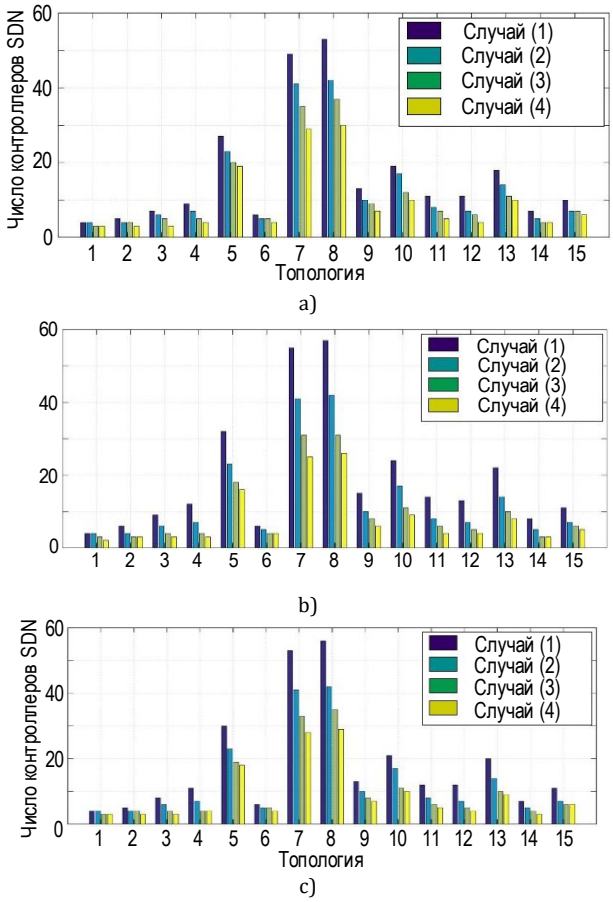


Рис. 2. Оптимальное число SDN-контроллеров для каждой топологии Сценария I (a), Сценария II (b), Сценария III (c)

Fig. 2. The Ideal Number of SDN Controllers for Each Topology in a Simulation Scenario I (a), Scenario II (b), Scenario III (c)

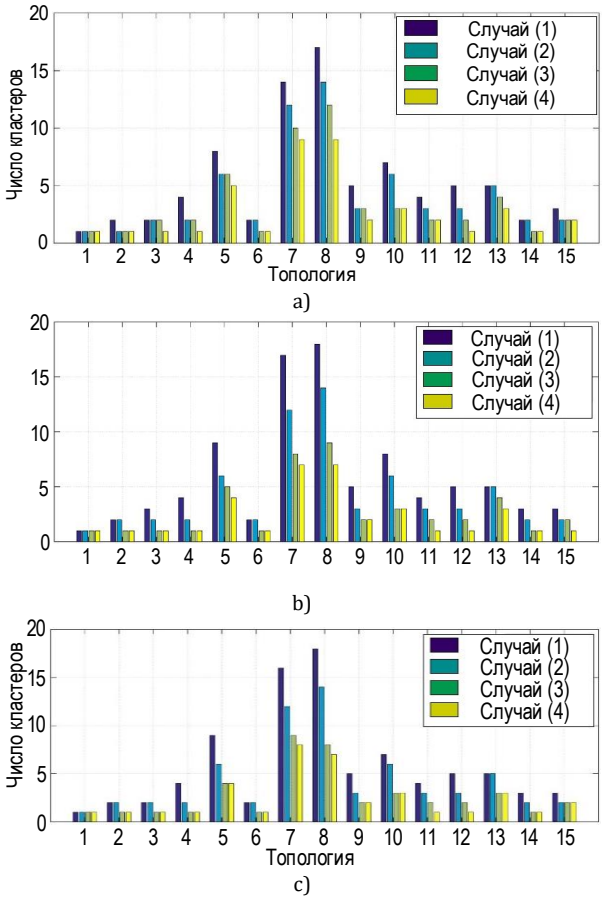


Рис. 3. Оптимальное число кластеров для каждой топологии Сценария I (a), Сценария II (b), Сценария III (c)

Fig. 3. The Ideal Number of Clusters for Each Topology in a Simulation Scenario I (a), Scenario II (b), Scenario III (c)

Для каждого сценария рассматривается десять значений каждого параметра, а каждое значение представляет собой случай моделирования. В таблице 4 приведены значения рассматриваемых параметров в каждом случае по каждому сценарию.

На рисунке 4 представлены результаты трех упомянутых в таблице 4 сценариев моделирования.

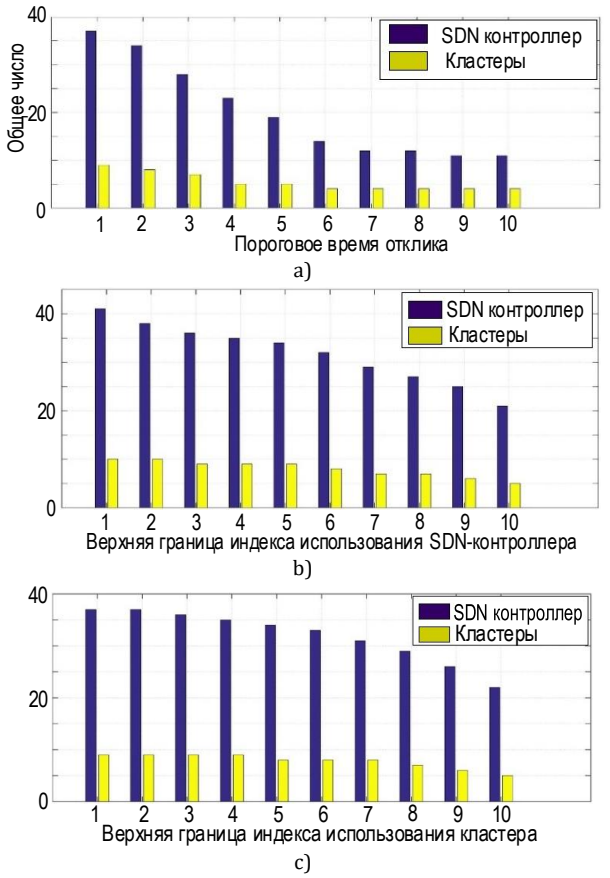


Рис. 4. Оптимальное число SDN-контроллеров и кластеров для Сценария I (a), Сценария II (b), Сценария III (c)

Fig. 4. The Ideal Number of SDN Clusters and Controllers for Simulated Scenario I (a), Scenario II (b), Scenario III (c)

ТАБЛИЦА 4. Десять наборов параметров для трех сценариев моделирования

TABLE 4. Ten Sets of Parameters for Three Simulation Scenarios

Ссылка. Номер	Сценарий I	Сценарий II	Сценарий III
Случай (1)	$T_{thr\text{-}1}$ = 1 мс	U_{ub1} = 0,86	$U_{C\text{-}ub1}$ = 0,86
Случай (2)	$T_{thr\text{-}2}$ = 2 мс	U_{ub2} = 0,87	$U_{C\text{-}ub2}$ = 0,87
Случай (3)	$T_{thr\text{-}3}$ = 3 мс	U_{ub3} = 0,88	$U_{C\text{-}ub3}$ = 0,88
Случай (4)	$T_{thr\text{-}4}$ = 4 мс	U_{ub4} = 0,89	$U_{C\text{-}ub4}$ = 0,89
Случай (5)	$T_{thr\text{-}5}$ = 5 мс	U_{ub5} = 0,90	$U_{C\text{-}ub5}$ = 0,90
Случай (6)	$T_{thr\text{-}6}$ = 6 мс	U_{ub6} = 0,91	$U_{C\text{-}ub6}$ = 0,91
Случай(7)	$T_{thr\text{-}7}$ = 7 мс	U_{ub7} = 0,92	$U_{C\text{-}ub7}$ = 0,92
Случай (8)	$T_{thr\text{-}8}$ = 8 мс	U_{ub8} = 0,93	$U_{C\text{-}ub8}$ = 0,93
Случай (9)	$T_{thr\text{-}9}$ = 9 мс	U_{ub9} = 0,94	$U_{C\text{-}ub9}$ = 0,94
Случай (10)	$T_{thr\text{-}10}$ = 10 мс	U_{ub10} = 0,95	$U_{C\text{-}ub10}$ = 0,95

Из рисунка 4a (Сценарий I) видно, что по мере увеличения порогового времени отклика SDN-контроллера общее количество оптимальных SDN-контроллеров и, соответственно, кластеров, необходимых для сети, уменьшается. Это аналогично Сценарию II (см. рисунок 4b) и Сценарию III (см. рисунок 4c). Однако влияние изменения порогового времени отклика на оптимальное количество SDN-контроллеров и кластеров значительнее, чем влияние индексов использования.

Разработанный CSSA сравнивается с генетическим алгоритмом (GA, *аббр. от англ.* Genetic Algorithm), алгоритмом PSO и алгоритмом оптимизации «серого волка» (GWO, *аббр. от англ.* Grey Wolf Optimization). Это самые последние алгоритмы, используемые для решения задачи размещения контроллеров в мультиконтроллерных сетях SDN. Они реализованы для ранее представленной случайной сети со 100 OpenFlow-коммутаторами, для указанных трех сценариев в таблице 4. Для такого сравнения рассматриваются две основные метрики: доля отказов в обслуживании со стороны контроллера и использование системы контроллеров в целом. Естественно, что при этом анализируются случаи для различных значений длительности задержки. На рисунках 5 и 6 представлено процент отказов и общее использование системы для каждого алгоритма.

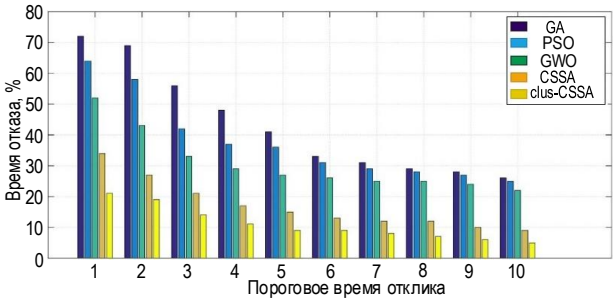


Рис. 5. Доля отказов в обслуживании со стороны контроллера при использовании сравниваемых алгоритмов

Fig. 5. The Percentage of Controller Denials of Service While Utilising Comparative Techniques

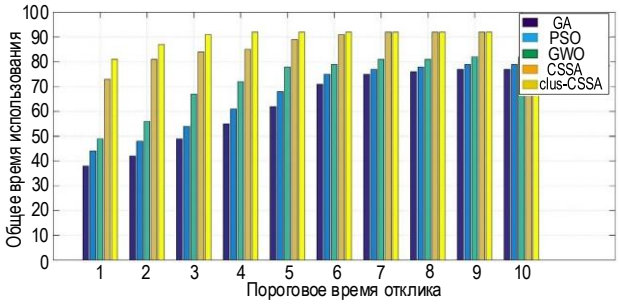


Рис. 6. Общее использование системы для сравниваемых алгоритмов

Fig. 6. System Usage in General for Comparing Methods

Результаты показывают, что разработанный оптимизированный CSSA достигает более высокой эффективности, чем другие алгоритмы.

Заключение

В работе представлено решение научной проблемы размещения контроллеров в мультиконтроллерных сетях и балансировки нагрузки, новизна которого (решения) состоит в следующем.

Во-первых, предложен метод построения мультиконтроллерной сети, основанный на интегральном решении задач по размещению контроллеров в таких сетях, базирующийся на метаэвристическом (вследствие сложности решаемых задач) алгоритме и алгоритме балансировки нагрузки, позволяющем обеспечить наилучшее использование ресурсов контроллеров. Во-вторых, предложено использовать иерархическую кластеризацию такой сети, включающую в себя кластеры с головными узлами и централизованный контроллер, что обеспечивает балансировку нагрузки в разработанном методе построения сети. В-третьих, разработан модифицированный алгоритм CSSA для использования в иерархических кластерных сетях clus-CSSA.

Разработанный метод построения мультиконтроллерной сети позволяет уменьшить долю отказов в обслуживании со стороны контроллера и увеличить общее использование системы во всем диапазоне изменения задержки от 1 до 10 мс по сравнению как с широко известными метаэвристическими алгоритмами PSO и GWO, так и с предыдущей версией CSSA. При этом для наиболее сложного случая задержки величиной в 1 мс выигрыш по доле отказов и по общему использованию системы достигает значения более, чем в 2 раза.

В дальнейшем планируется реализовать алгоритм CSSA для сети Интернета Вещей высокой плотности.

Список источников

1. Кучерявый А.Е., Маколкина М.А., Киричек Р.В. Тактильный интернет. Сети связи со сверхмалыми задержками // Электросвязь. 2016. № 1. С. 44–46.

2. Бородин А.С., Кучерявый А.Е. Сети связи пятого поколения как основа цифровой экономики //Электросвязь. 2017. № 5. С. 45–49.

3. Кучерявый А.Е., Прокопьев А.В., Кучерявый Е.А. Самоорганизующиеся сети. СПб: Типография «Любавич», 2011. 312 с.

4. Атея А.А., Мутханна А.С., Кучерявый А.Е. Интеллектуальное ядро для сетей связи 5G и тактильного интернета на базе программно-конфигурируемых сетей // Электросвязь. 2019. № 3. С. 34–40.

5. Heller B., Sherwood R., McKeown N. The controller placement problem // Proceedings of the Special Interest Group on Data Communication (SIGCOMM '12, Helsinki, Finland, 13 August–17 August 2012). Special October issue ACM SIGCOMM Computer Communication Review. New York: ACM, 2012. Vol. 42. Iss. 4. PP. 473–478. DOI:10.1145/2377677. 2377767

6. Yao G., Bi J., Li Y., Guo L. On the Capacitated Controller Placement Problem in Software Defined Networks // IEEE Communications Letters. 2014. Vol. 18. Iss. 8. PP. 1339–1342. DOI:10.1109/LCOMM.2014.2332341

7. Dixit A., Hao F., Mukherjee S., Lakshmanet T.V., Kompella R. Towards an elastic distributed SDN controller // Proceedings of the Special Interest Group on Data Communication (SIGCOMM '13, Hong Kong, China, 16 August 2013). Special October issue ACM SIGCOMM Computer Communication Review. New York: ACM, 2013. Vol. 43. Iss. 4. PP. 7–12. DOI: 10.1145/2534169. 2491193

8. Ozsoy F.A., Pinar M.C. An exact algorithm for the capacitated vertex pcenter problem // Computers & Operations Research. 2006. Vol. 33. Iss. 5. PP. 1420–1436. DOI:10.1016/j.cor.2004.09.035

9. Sahoo K.S., Sarkar A, Mishra S.K., Sahoo B., Puthal D., Obaidat M.S., et al. Metaheuristic Solutions for Solving Controller Placement Problem in SDN-based WAN Architecture // Proceedings of the 14th International Joint Conference on e-Business and Telecommunications (ICETE 2017) and 8th International Conference on Data Communication Networking (DCNET), Madrid, Spain, 15–23 July 2017. SciTePress Digital Library, 2017. PP. 15–23. DOI:10.5220/0006483200150023

10. Rath H.K., Revoori V., Nadaf S.M., Simha A. Optimal controller placement in Software Defined Networks (SDN) using a non-zero-sum game // Proceedings of the International Symposium on a World of Wireless, Mobile and Multimedia Networks (Sydney, Australia, 19 June 2014). IEEE, 2014. DOI:10.1109/WoWMoM.2014.6918987

11. Ksentini A., Bagaa M., Taleb T., Balasingham I. On using bargaining game for Optimal Placement of SDN controllers // Proceedings of the International Conference on Communications (ICC, Kuala Lumpur, Malaysia, 22–27 May 2016). IEEE, 2016. DOI:10.1109/ICC.2016.7511136

12. Ateya A.A., Muthanna A., Vybornova A., Algarni A.D., Abuarqoub A., Koucheryavy Y., et al. Chaotic salp swarm algorithm for SDN multi-controller networks // Engineering Science and Technology, an International Journal. 2019. Vol. 22. Iss. 4. PP. 1001–1012. DOI:10.1016/j.jestch.2018.12.015

13. Killi B.P., Rao S.V. Capacitated Next Controller Placement in Software Defined Networks // IEEE Transactions on Network and Service Management. 2017. Vol. 14. Iss. 3. PP. 514–527. DOI:10.1109/TNSM.2017.2720699

14. Chen W, Chen C, Jiang X, Liu L. Multi-Controller Placement Towards SDN Based on Louvain Heuristic Algorithm // IEEE Access. 2018. Vol. 6. PP. 49486–49497. DOI:10.1109/ACCESS.2018.2867931

15. Wang G, Zhao Y, Huang J, Duan Q., Li J. A K-means-based network partition algorithm for controller placement in software defined network // Proceedings of the International Conference on Communications (ICC, Kuala Lumpur, Malaysia, 22–27 May 2016). IEEE, 2016. DOI:10.1109/ICC.2016.7511441

16. Kuang H., Qiu Y., Li R., Liu X. A Hierarchical K-Means Algorithm for Controller Placement in SDN-Based WAN Architecture // Proceedings of the 10th International Conference on Measuring Technology and Mechatronics Automation (ICMTMA, Changsha, China, 10–11 February 2018). IEEE, 2018. PP. 263–267. DOI:10.1109/ICMTMA.2018.00070

17. Hu Y., Wang W., Gong X., Que X., Cheng S. BalanceFlow: Controller load balancing for OpenFlow networks // Proceedings of the 2nd International Conference on Cloud Computing and Intelligent Systems (Hangzhou, China, 30 October 2012–01 November 2012. IEEE, 2013. PP. 780–785. DOI:10.1109/CCIS.2012.6664282

References

1. Koucheryavy A.E., Makolkina M.A., Kirichek R.V. Tactile internet. ULTRA-Low Latency Networks. *Electrosvyaz*. 2016;1:44–46. (in Russ.)

2. Borodin A.S., Koucheryavy A.E. Fifth generation networks as a base to the digital economy. *Electrosvyaz*. 2017;5(45–49).

3. Koucheryavy A.E., Prokopiev A.V., Koucheryavy Y.A. *Self-Organizing Network*. St. Petersburg: Lyubavich Printing House; 2011. 312 p. (in Russ.)

4. Ateya A.A., Muthanna A.S., Koucheryavy A.E. Intelligent core network for 5g and tactile internet systems based on software defined networks. *Electrosvyaz*. 2019;3:34–40. (in Russ.)

5. Heller B., Sherwood R., McKeown N. The controller placement problem. *Proceedings of the Special Interest Group on Data Communication, SIGCOMM '12, 13 August–17 August 2012, Helsinki, Finland. Special October issue ACM SIGCOMM Computer Communication Review*. New York: ACM; 2012;42(4):473–478. DOI:10.1145/2377677.2377767

6. Yao G., Bi J., Li Y., Guo L. On the Capacitated Controller Placement Problem in Software Defined Networks. *IEEE Communications Letters*. 2014;18(8):1339–1342. DOI:10.1109/LCOMM.2014.2332341

7. Dixit A., Hao F., Mukherjee S., Lakshmanet T.V., Kompella R. Towards an elastic distributed SDN controller. *Proceedings of the Special Interest Group on Data Communication, SIGCOMM '13, 16 August 2013, Hong Kong, China. Special October issue ACM SIGCOMM Computer Communication Review*. New York: ACM; 2013;43(4):7–12. DOI: 10.1145/2534169.2491193

8. Ozsoy F.A., Pinar M.C. An exact algorithm for the capacitated vertex pcenter problem. *Computers & Operations Research*. 2006;33(5):1420–1436. DOI:10.1016/j.cor.2004.09.035

9. Sahoo K.S., Sarkar A, Mishra S.K., Sahoo B., Puthal D., Obaidat M.S., et al. Metaheuristic Solutions for Solving Controller Placement Problem in SDN-based WAN Architecture. *Proceedings of the 14th International Joint Conference on e-Business and Telecommunications (ICETE 2017) and 8th International Conference on Data Communication Networking (DCNET), Madrid, Spain, 15–23 July 2017*. SciTePress Digital Library; 2017. p.15–23. DOI:10.5220/0006483200150023

10. Rath H.K., Revoori V., Nadaf S.M., Simha A. Optimal controller placement in Software Defined Networks (SDN) using a non-zero-sum game. *Proceedings of the International Symposium on a World of Wireless, Mobile and Multimedia Networks, 19 June 2014, Sydney, Australia*. IEEE; 2014. DOI:10.1109/WoWMoM.2014.6918987

11. Ksentini A., Bagaa M., Taleb T., Balasingham I. On using bargaining game for Optimal Placement of SDN controllers. *Proceedings of the International Conference on Communications, ICC, 22–27 May 2016, Kuala Lumpur, Malaysia*. IEEE; 2016. DOI:10.1109/ICC.2016.7511136

12. Ateya A.A., Muthanna A., Vybornova A., Algarni A.D., Abuarqoub A., Koucheryavy Y., et al. Chaotic salp swarm algorithm for SDN multi-controller networks. *Engineering Science and Technology, an International Journal*. 2019;22(4):1001–1012. DOI:10.1016/j.jestch.2018.12.015

13. Killi B.P., Rao S.V. Capacitated Next Controller Placement in Software Defined Networks. *IEEE Transactions on Network and Service Management.* 2017;14(3):514–527. DOI:10.1109/TNSM.2017.2720699

14. Chen W, Chen C, Jiang X, Liu L. Multi-Controller Placement Towards SDN Based on Louvain Heuristic Algorithm. *IEEE Access*. 2018;6:49486–49497. DOI:10.1109/ACCESS.2018.2867931

15. Wang G, Zhao Y, Huang J, Duan Q., Li J. A K-means-based network partition algorithm for controller placement in software defined network. *Proceedings of the International Conference on Communications, ICC, 22–27 May 2016, Kuala Lumpur, Malaysia*. IEEE; 2016. DOI:10.1109/ICC.2016.7511441

16. Kuang H., Qiu Y., Li R., Liu X. A Hierarchical K-Means Algorithm for Controller Placement in SDN-Based WAN Architecture. *Proceedings of the 10th International Conference on Measuring Technology and Mechatronics Automation, ICMTMA, 10–11 February 2018, Changsha, China*. IEEE; 2018. p.263–267. DOI:10.1109/ICMTMA.2018.00070

17. Hu Y., Wang W., Gong X., Que X., Cheng S. BalanceFlow: Controller load balancing for OpenFlow networks. *Proceedings of the 2nd International Conference on Cloud Computing and Intelligent Systems, 30 October 2012–01 November 2012, Hangzhou, China*. IEEE; 2013. p.780–785. DOI:10.1109/CCIS.2012.6664282

Статья поступила в редакцию 02.05.2023; одобрена после рецензирования 10.05.2023; принята к публикации 17.05.2023.
The article was submitted 02.05.2023; approved after reviewing 10.05.2023; accepted for publication 17.05.2023.

Информация об авторе:

МУТХАННА Аммар Салех Али

кандидат технических наук, доцент кафедры сети связи и передачи данных Санкт-Петербургского государственного университета телекоммуникаций им. проф. М.А. Бонч-Бруевича

https://orcid.org/0000-0003-0213-8145

DOI:10.31854/1813-324X-2023-9-2

Научная статья
УДК 004.42
DOI:10.31854/1813-324X-2023-9-2-95-111

Моделирование программы с уязвимостями с позиции эволюции ее представлений. Часть 2. Аналитическая модель и эксперимент

Константин Евгеньевич Израилов, konstantin.izrailov@mail.ru

Санкт-Петербургский Федеральный исследовательский центр Российской академии наук,
Санкт-Петербург, 199178, Российская Федерация

Аннотация: *Изложены результаты исследования процесса создания программ и возникающих при этом уязвимостей. Во второй части цикла статей предлагается обобщенная аналитическая модель жизненного цикла программы на базе ее представлений, учитывающая способы прямого и обратного преобразования последних. Также в модели отражается возникновение и обнаружения уязвимостей и их классификация. Из нее синтезируется частная модель, отражающая текущее состояние эволюции программных представлений, и исходя из которой был выведен ряд основополагающих утверждений, записанных в аналитическом виде. Для основания работоспособности моделей в части отражения уязвимостей проводятся два следующих эксперимента: ретроспективно-фактологический, сопоставляющий реально существующие уязвимости с частной моделью; и практический, демонстрирующий эволюцию уязвимости в представлениях в процессе эволюции простейшей программы. В результате второго эксперимента наглядно показан рост охвата уязвимостью представлений в процессе разработки.*

Ключевые слова: *программная инженерия, информационная безопасность, уязвимости, представления программы, жизненный цикл, моделирование*

Ссылка для цитирования: Израилов К.Е. Моделирование программы с уязвимостями с позиции эволюции ее представлений. Часть 2. Аналитическая модель и эксперимент // Труды учебных заведений связи. 2023. Т. 9. № 2. С. 95–111. DOI:10.31854/1813-324X-2023-9-2-95-111

Modeling a Program with Vulnerabilities in the Terms of the Its Representations Evolution. Part 2. Analytical Model and Experiment

Konstantin Izrailov, konstantin.izrailov@mail.ru

Saint-Petersburg Federal Research Center of the Russian Academy of Sciences,
St. Petersburg, 199178, Russian Federation

Abstract: *The investigation results of the creating programs process and the resulting vulnerabilities are presented. In the second part of the articles series, a program life cycle generalized analytical model of the based on its representations is proposed, taking into account the direct and reverse transformation methods. Also, the model reflects the occurrence and detection of vulnerabilities and their classification. A particularly model is synthesized from it, reflecting the current state of the program representations evolution, and on the basis of which a number of fundamental statements were derived, written in an analytical form. To base the performance of models in*

reflecting vulnerabilities terms, the following two experiments are carried out: retrospective-factual, comparing real-life vulnerabilities with a particularly model; and practical demonstration the evolution of vulnerability in representations in the process of the simplest program evolution. As a result of the second experiment, the increase in coverage by the vulnerability of representations in the development process is clearly shown.

Keywords: *software engineering, information security, vulnerabilities, program representations, life cycle, modeling*

For citation: Izrailov K. Modeling a Program with Vulnerabilities in the Terms of Its Representations Evolution. Part 2. Analytical Model and Experiment. *Proc. of Telecom. Universities.* 2023;9(2):95–111. (in Russ.) DOI:10.31854/1813-324X-2023-9-2-95-111

1. ВВЕДЕНИЕ

Поиск уязвимостей в программном обеспечении является актуальнейшей задачей в сфере информационной безопасности (далее – ИБ). Наличие уязвимостей в зависимости от области функционирования программ может приводить к угрозам различной степени тяжести, вплоть до человеческих жертв (что особенно актуально для киберфизических устройств [1–3]). Однако, несмотря на достаточно долгую историю решения данной задачи, многие успехи в обнаружении и нейтрализации уязвимостей нивелируются новыми способами их внедрения злоумышленниками. Причина этого, в том числе, может заключаться в слабой формализации не только самого понятия уязвимости, но и всего жизненного цикла программы – от появления самой ее задумки до получения непосредственно выполняемого образца.

Часть важнейших и используемых далее результатов была получена в предыдущей части данного цикла статей [4]. Ранее автором были обоснованы и введены *представления* (далее – Представление), через которые проходит любая программа, а именно, следующие (с их краткой сутью):

1) Идея – изначальный замысел программы;

2) Концептуальная модель – основные понятия, их взаимосвязи и особенности программы;

3) Архитектура – структура программы с делением на физические и логические модули;

4) Двухмерная структурная схема – совмещение Представления алгоритмов в графическом и текстовом виде (например, язык ДРАКОН [5, 6]);

5) Функциональная диаграмма – аналог Двухмерной структурной схемы, но с некоторыми особенностями (например, FBD [7], SFC [8] или LD [9]);

6) Блок-схема – классическое блочное Представление алгоритма, как правило, в соответствии с ГОСТ 19.701-90 Единая система программной документации. Схемы алгоритмов, программ, данных и систем. Обозначения условные и правила выполнения;

7) Структурограмма – аналог Блок-схем, но для программ согласно структурной парадигме программирования; как следствие, без операций безусловного перехода (например, диаграммы Насси – Шнейдермана [10]);

8) Псевдокод – запись алгоритма в специальной текстовой нотации (например, семейство языков Algol [11]);

9) Классический исходный код – код на классических языках программирования (Pascal, Java и C/C++/C# и т. п.);

10) Метакод генерации – промежуточный код для генерации Классического исходного (например, формальные грамматики для генератора синтаксического анализатора Yacc/Bison [12]);

11) Сценарный код – развитие Классического кода в сторону повышения абстракции, упрощения разработки и укрупнения используемых элементов (сценарий командной строки Shell Script [13]);

12) Ассемблерный код – низкоуровневое Представление кода, отражающее специфику среды выполнения (например, согласно синтаксису для утилиты-ассемблера TASM [14]);

13) Дерево абстрактного синтаксиса – промежуточное Представление между человеко- и машиноориентированным кодом (так, Представление используется инструментом Bandit [15] для поиска уязвимостей коде Python-программ);

14) Машинный код – код программы, готовый для выполнения на Центральном процессоре (например, бинарный образ для UEFI-прошивка персонального или серверного компьютера [16]);

15) Байт-код – аналогичный Машинному коду, но предназначенный для выполнения на виртуальной машине (например, CIL (*аббр. от англ.* Common Intermediate Language – общий промежуточный язык) для .Net [17] и JBC (*аббр. от англ.* Java Byte Code) для Java [18]).

Для определения Представлений использовались понятия из категориальный пары – *форма* и *содержания* [19]. Первый элемент пары отвечает за внешнее отражение Представления; второй же соответствует внутреннему функционалу программы в данной форме (отражая тем самым первоначальную Идею программы); при этом, содержание само состоит из двух компонент – своей функциональной логики и базиса, на котором эта логика построена.

На базе Представлений была предложена схема жизненного цикла программы (далее – Схема), приведенная на рисунке 1.

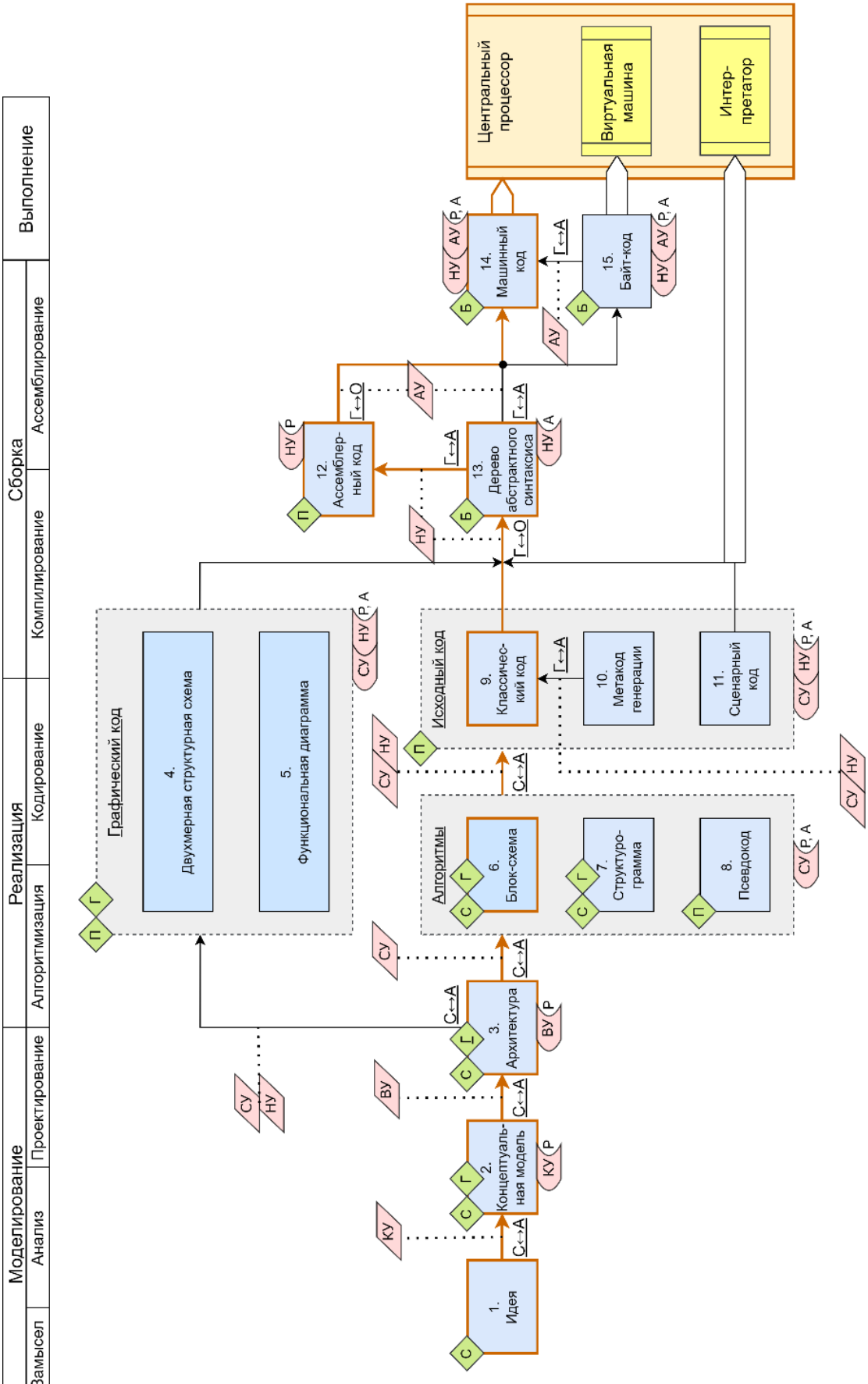


Рис. 1. Схема жизненного цикла Представлений программы с уязвимостями

Fig. 1. Life Cycle Scheme of Program Representation with Vulnerabilities

На Схеме используются следующие обозначения: таблицы сверху – стадии разработки программы и их состав; прямоугольник с синим фоном – само Представление; ромб с зеленым фоном – его форма (по первой букве): Словесная, Графическая, Программно-языковая и Бинарная; стрелки – способ получения Представления из предыдущего; надпись под стрелкой в формате «$X \leftrightarrow Y$» – тип прямого и обратного преобразования (по первой букве): Анализ, Синтез, Генерация, Обратная генерация; параллелограмм с красным фоном, соединенный пунктирной линией со способом получения – класс возникающей уязвимости согласно структурному уровню (по первой букве): Концептуальные, Высокоуровневые, Среднеуровневые, Низкоуровневые, Атомарные; прямоугольник с розовым фоном и гнутыми боками – класс уязвимости, обнаруживаемый в Представлении; буква внутри вогнутой части прямоугольника – тип способа обнаружения (по первой букве): Ручной, Автоматический; прямоугольник с боковыми линиями и желтым фоном – механизм выполнения конечного Представления программы.

Как результат, удалось ввести более формализованное и точное (с точки зрения автора) определение понятия «уязвимость», как *различие содержания программы в ее конечном и изначальном Представлениях*.

Классификация уязвимостей детализирована: по структурному уровню – согласно уровню абстракции, на котором внесена уязвимость (что отображается на Схеме); по изменению содержания – потеря, внесение и модификация функционала; по воздействию на внутренние информационные потоки (по аналогии с триадой нарушения конфиденциальности, целостности и доступности): появление, изменение и потеря потока.

А поскольку визуально Схема обладает признаками формализации (строго выделенные элементы, логика их связей, перечислимый набор свойств и т. п.), то целесообразно ее записать в виде соответствующей аналитической модели, а именно – жизненного цикла программы (далее – Модель); это позволит применять соответствующие математические аппараты, что будет иметь как теоретическую значимость – например, для доказательства решения различных задач по преобразованию Представлений, так и практическую значимость – для реализации программных средств управления комплексным методом (де)эволюции Представлений. Учет же в аналитической Модели классификации уязвимостей позволит ее применить для сферы ИБ программной инженерии.

2. ОБОБЩЕННАЯ АНАЛИТИЧЕСКАЯ МОДЕЛЬ

В интересах строгой записи аналитической Модели введем следующие обозначения (строчными буквами) упоминаемых ранее основополагающих объектов предметной области:

– p (*от англ.* Presentation) – Представление;
– f (*от англ.* Form) – форма;
– c (*от англ.* Content) – содержание;
– v (*от англ.* Vulnerability) – уязвимость.

Подчеркнем, что все объекты предметной области так или иначе связаны с сущностями информационного мира.

Следуя той же логике, введем обозначения (прописными буквами) операций над объектами, детально описанные в предыдущей части цикла статей:

– T (*от англ.* Transformation) – преобразование Представления некоторым способом;
– $\overline{T}$ (верхняя черта соответствует отрицанию) – обратное преобразование Представления некоторым способом;
– I (*от англ.* Infection) – внесение уязвимости некоторым способом;
– D (*от англ.* Detection) – обнаружение уязвимости некоторым способом;
– E – пустое действие (т. е. отсутствие операции, как таковой).

В результате действий получаются новые объекты (в том числе, тождественные исходным).

Также будем использовать следующие достаточно распространенные (и, что важно, интуитивно понятные и имеющие лаконичную запись) обозначения математического аппарата:

1) $\circ$ – действие над объектом (аналог функции от одного операнда, что позволяет записывать последовательность операций в хорошо читаемом виде);

2) $\langle x|y\rangle$ – совокупность объектов или операций x и y (используется для разделения на форм-содержательные аспекты более сложной сущности);

3) правый верхний индекс – обозначение номера Представления (индекс будет отсутствовать тогда, когда отнесение объекта или процесса к конкретному Представлению не существенно); индексы тождественны номерам Представлений на Схеме; исключением является формат индекса «$x \to y$», который подчеркивает, что идет преобразование Представления с индексом x к индексу y.

4) используются следующие наименования индексов: k – для Представлений; i – для перечисления объектов произвольного рода; x и y – произвольные для дополнительных целей; a, n и m – для примеров;

5) $\equiv$ – тождественность выражений (т. е. верность при любых значениях их переменных);

6) $\setminus$ – отличие между объектами (аналог вычитания множеств);

7) $x \to y$ – некоторый алгоритм по преобразования объекта x к объекту y;

8) $\triangleright$ – специальное преобразование, которое согласованно изменяет логику и базис содержания (т. е. после применения изменяется не все содержание, а его компоненты);

9) $\emptyset$ – несуществующий объект (может относиться к Представлению или уязвимости);

10) $\mathbb{C}$ – получение класса объекта (может состоять из подклассов, обозначаемых в виде подписей в нижней правой части символа).

11) $\cdot$ – оператор составления нового объекта из подобъектов (используется для повышения читаемость записи);

12) $\hat{x}$ («шапка» над символом) – пометка того, что объект содержит уязвимость (используется контекстно для улучшения восприятия записи).

Далее для упрощения будем Представление с индексом k называть просто – Представление k.

Исходя из введенных обозначений и используя сделанные ранее утверждения, а также продуцируя новые, Схема в аналитическом виде может быть записана следующим образом.

Утверждение 1. Представление является совокупностью формы и содержания:

$$p^k \equiv \langle f|c\rangle^k \equiv \langle f^k|c^k\rangle.$$

Иная, более интуитивная, совокупность формы и содержания следующая:

$$\langle f|c\rangle \equiv \boxed{f\boxed{c}} \equiv c \underline{by} f,$$

где $\boxed{\quad}$ и $\underline{by}$ – «представимость» (содержания с в форме f); впрочем, данная форма далее не применяется, а указана лишь для расширения визуальной части предлагаемого математического аппарата.

Введем также класс формы, а именно следующий:

$$\mathbb{C}_{form}, form \in \{VF, GF, PF, BF\},$$

где $form$ соответствует одной из введенных ранее форм: словесной (VF), графической (GF), программной (PF) и бинарной (BF), где первая буква является *сокр. от англ.* названия (Verbal, Graphic, Program, Binary), а вторая – *сокр. от англ.* Form.

Утверждение 2. Исходным Представлением любой программы является идея:

$$p^0 \equiv \langle f|c\rangle^0 \equiv \langle f^0|c^0\rangle,$$

где 0 – индекс Представление Идеи.

Утверждение 3. Форма и содержания отражают программу в некотором одном Представлении и не могут составлять совокупность, находясь в разных Представлениях:

$$(\forall x, y: x \neq y) \Rightarrow \langle f^x|c^y\rangle \in \emptyset,$$

где $\in \emptyset$ означает, что объект (т. е. Представление $\langle f^x|c^y\rangle$) не существует.

Другими словами, содержание Представления может сосуществовать только с формой этого же Представления. Для обоснования этого можно привести следующий пример – логика работы Машинного кода не может быть отражена в форме Исходного кода, поскольку первый оперирует инструкциями процессора (ориентированными на автомат), а второй – конструкциями языка программирования (ориентированными на человека).

Утверждение 4. Жизненный цикл программы состоит из преобразования Представлений:

$$p^{k+1} = T^{k\to k+1} \circ p^k,$$

где $k+1$ – индекс нового Представления, получаемого из Представления k. Индекс $k \to k+1$ означает, что способ преобразования предназначен для оперирования с формой f^k и содержанием c^k с получением аналогичных для индекса $k+1$.

Саму операцию преобразования (как и любую другую, применяемую к Представлению) можно разбить на 2 части, алгоритмы которых работают с формой и содержанием:

$$\begin{cases} T^{k\to k+1} \equiv \langle (T^{k\to k+1})_f | (T^{k\to k+1})_c \rangle \\ (T^{k\to k+1})_f \equiv f^k \to f^{k+1} \\ (T^{k\to k+1})_c \equiv c^k \triangleright c^{k+1} \end{cases},$$

где $(T^{k\to k+1})_f$ – изменение формы Представления k к $k+1$; $(T^{k\to k+1})_c$ – перевод логики содержания на базисе Представления k к логике Представления на базисе $k+1$; $\to$ – алгоритм изменения формы; $\triangleright$ – алгоритм перевода логики на другой базис.

Подчеркнем, что поскольку преобразование происходит без внесения уязвимостей, то само содержание не меняется, а осуществляется лишь согласованное изменение его логики и базиса; например, при кодировании Алгоритмов с помощью Исходного кода меняется не само функционирование, а логика и объекты, на которых она строится – например, с элементов блок-схем к конструкциям языка программирования.

Введем классификацию преобразований согласно веденной ранее:

$$\mathbb{C}_{transform}, transform \in \{ST, GT\},$$

где $transform$ соответствует одному из введенных ранее типов прямого преобразования: синтез (ST) и генерация (GT); первая буква названия является *сокр. от англ.* названия (Synthes, Generation), а вторая – *сокр. от англ.* Transform, *перев. на рус.* Трансформировать.

Утверждение 5. Уязвимость Представления есть отличие между содержаниями Представлений Идеи и текущего (независимо от их формы):

$$v \equiv c^0 \setminus c^k,$$

где k – индекс текущего Представления.

Утверждение 6. Отсутствие уязвимости соответствует тождественности Представлений Идеи и текущего:

$$c^0 \equiv c^k \Leftrightarrow c^0 \setminus c^k = v \in \emptyset,$$

где $v \in \emptyset$ означает, что какая-либо уязвимость не существует.

Утверждение 7. Классификация уязвимости возможна по ее структурному уровню, изменению содержаний Представлений и воздействию на информационные потоки:

$$\begin{cases} \mathbb{C}_v \equiv (\mathbb{C}_v)_{level} \cdot (\mathbb{C}_v)_{difference} \cdot (\mathbb{C}_v)_{infoflow} \\ level \in \{CL, HL, ML, LL, AL\} \\ difference \in \{-, +, \triangle\} \\ infoflow \in \{-, +, \triangle\} \end{cases},$$

где $\mathbb{C}_v$ – класс уязвимости (составной); $(\mathbb{C}_v)_{level}$ – подклассы согласно структурному уровню $level$: КУ (CL), ВУ (HL), СУ (ML), НУ (LL), АУ (AL), где первая буква названия является *сокр. от англ.* аналогов названий уязвимостей (Conceptual, High, Middle, Low, Atomic), а вторая – сокращение переведенного *на англ. яз.* слова «уровень» (Level); $(\mathbb{C}_v)_{difference}$ – подклассы согласно изменению в содержании $difference$: потеря функционала ($-$), внесение функционала ($+$), модификация функционала ($\triangle$); $(\mathbb{C}_v)_{infoflow}$ – подклассы согласно изменению информационных потоков $infoflow$ (*сокр. от* Information Flow): потеря информационного потока ($-$), возникновение информационного потока ($+$), изменение информационного потока ($\triangle$).

Так, если обнаружена уязвимость в Алгоритме системы аутентификации (что соответствует СУ), реализованная с помощью добавление пароля по умолчанию (соответствует внесению функционала) и позволяющая проходить аутентификацию несанкционированным для этого объектам-пользователям (соответствует появлению информационного потока), то ее составной класс будет записан следующим образом:

$$\mathbb{C}_v = \mathbb{C}_{ML++}.$$

Утверждение 8. Уязвимость может быть внесена операцией, которая меняет содержания в некотором Представлении:

$$\hat{p} \equiv \widehat{\langle f|c\rangle} \equiv \langle f|\hat{c}\rangle = I \circ p,$$

где $\hat{p}$ или тождественное ему $\widehat{\langle f|c\rangle}$ – Представление с внесенной уязвимостью (т. е. уже небезопасное); $\hat{c}$ – содержание Представления с этой же уязвимостью (форма, очевидно, сама по себе из-за внесения уязвимости не меняется, поскольку также используется для отображения уязвимости).

Утверждение 9. Операция внесения уязвимости может быть записана, как способ изменения содержания, при котором форма Представления является лишь связывающим звеном (программы и способа) и на уязвимость никак не влияет:

$$\begin{cases} \mathrm{I}^k \equiv \langle (E)_f | (I^k)_c \rangle \\ (I^k)_c \equiv c^k \to \hat{c}^k \end{cases},$$

где $(E)_f$ – отсутствие изменений в конкретной форме Представления; $(I^k)_c$ – изменение содержания Представления k при внесении уязвимостей; $\hat{c}^k$ – содержание с внесенной уязвимостью; $\to$ – некий алгоритм изменения содержания (также отразится на подклассе уязвимости $(\mathbb{C}_v)_{difference}$).

Утверждение 10. Уязвимость может быть найдена путем применения операции обнаружения к Представлению:

$$v = D \circ p$$

Естественно, не каждое Представление подходит для обнаружения всех классов уязвимостей (например, «просчеты» в концептуальной модели практически невозможно найти по Машинному коду), о чем будет дано следующее утверждение. При этом, далеко не все уязвимости могут быть найдены автоматически и требуют применения ручного труда эксперта; для учета этого можно ввести классификацию способов поиска:

$$\mathbb{C}_{detect}, detect \in \{HD, AD\},$$

где $detect$ соответствует ручной (HD) или автоматической (AD) форме, где первая буква названия является *сокр. от англ.* названия (Hand, Auto), а вторая – *сокр. от англ.* Detect, *перев. на рус.* Обнаружить.

Утверждение 11. Способы обнаружения уязвимостей взаимодействуют с содержанием посредством формы, и, следовательно, могут применяться (а точнее иметь не 0-й эффект) только на соответствующем Представлении:

$$(\forall x, y: x \neq y) \Rightarrow D^x \circ p^y \in \emptyset,$$

где $\in \emptyset$ означает, что по данному Представлению данным способом не будет найдено ни одной уязвимости.

А поскольку (согласно Утверждению 5) суть операции обнаружения состоит в определении отличий между содержаниями двух Представлений, и при этом (согласно текущему Утверждению) обнаружение действует только для того Представления, для которого был создан его способ, то логично возникает следующее (производное) утверждение.

Утверждение 12. Операция обнаружения уязвимости может быть записана, как способ нахождения отличия содержания Представлений Идеи от содержания текущего Представления, при котором форма Представления является лишь связывающим звеном (программы и способа) и на отличие не влияет:

$$\begin{cases} \mathrm{D}^k \equiv \langle (E)_f | (D^k)_c \rangle \\ (D^k)_c \equiv c^0 \setminus c^k \end{cases},$$

где $(E)_f$ – не учет способом обнаружения конкретной формы Представления; $(D^k)_c$ – применение способа обнаружения к Представлению только в части содержания.

Утверждение 13. Восстановление Предыдущих Представлений из текущего согласно жизненному циклу программы состоит из обратных преобразований Представлений:

$$p^{k-1} = \bar{T}^{k \to k-1} \circ p^k,$$

где $k-1$ – индекс предыдущего Представления, получаемого из текущего Представления k; индекс $k \to k-1$ означает, что способ обратного преобразования предназначен для оперирования с формой f^k и содержанием c^k, получая аналогичные для индекса $k-1$.

По аналогии с прямым преобразованием (согласно Утверждению 4) операция обратного преобразования разбивается на 2 части, алгоритмы которых работают с формой и содержанием:

$$\begin{cases} \bar{\mathrm{T}}^{k \to k-1} \equiv \langle (T^{k \to k-1})_f | (T^{k \to k-1})_c \rangle \\ (T^{k \to k-1})_f \equiv f^k \to f^{k-1} \\ (T^{k \to k-1})_c \equiv c^k \triangleright c^{k-1} \end{cases},$$

где $(T^{k \to k-1})_f$ – изменение формы Представления k к Представлению $k-1$; $(T^{k \to k-1})_c$ – перевод логики содержания на базисе Представления k к логике на базисе Представления $k-1$.

Аналогично прямому преобразованию, в процессе операции уязвимости не появляются.

Введем классификацию преобразований согласно указанной ранее:

$$\mathbb{C}_{\overline{transform}}, \overline{transform} \in \{AT, RT\},$$

где $\overline{transform}$ соответствует одному из введенных ранее типов обратного преобразования: анализ (AT) и обратная генерация (RT); первая буква названия является *сокр. от англ.* Analysis, Reverse generation, а вторая – *сокр. от англ.* Transform.

Данный процесс, называемый *реверс-инжинирингом*, как раз и призван частично решить основную проблему поиска уязвимостей, внесенных на более ранних Представлениях, чем исследуемое.

Утверждение 14. Весь жизненный цикл программы без внесения уязвимостей может быть представлен, как множество цепочек преобразований Представлений:

$$\langle f^n | c^n \rangle = p^n = T^{n-1 \to n} \circ p^{n-1} = \\ = T^{n-1 \to n} \circ \ldots \circ T^{0 \to 1} \circ \langle f^0 | c^0 \rangle.$$

Проверим корректность действия операции обнаружения уязвимостей для такого случая безопасной (с позиции ИБ) программной инженерии. Поскольку в процессе создания программы уязвимости не вносились, то (согласно Утверждению 6) $c^0 \equiv c^n$. Тогда результат попытки обнаружения уязвимостей в конечном Представлении (согласно Утверждениям 10 и 12) будет следующим:

$$(c^0 \equiv c^n) \Longrightarrow v^n = D^k \circ \langle f^n | c^n \rangle = \langle f^n | c^0 \setminus c^n \rangle = \\ = \langle f^n | \emptyset \rangle \in \emptyset, \text{ т. е. } v \in \emptyset.$$

Таким образом, формальное применение операции обнаружения уязвимостей для безопасного Представления программы не позволило получить какую-либо уязвимость, что подтверждает тождественность сделанных утверждений и их аналитических записей.

Расширим аналитическую запись цепочки жизненного цикла программы для случая наличия уязвимостей.

Утверждение 15. Весь жизненный цикл программы с внесением уязвимостей (для общности, на каждом преобразовании) может быть представлен, как множество цепочек преобразований Представлений:

$$\langle \widehat{f^n | c^n} \rangle = I^n \circ \hat{p}^n = I^n \circ T^{n-1 \to n} \circ \hat{p}^{n-1} = \\ = I^n \circ T^{n-1 \to n} \circ I^{n-1} \circ \ldots \circ I^1 \circ T^{0 \to 1} \circ \langle f^0 | c^0 \rangle.$$

Применим описанные элементы аналитической Модели к реальным ситуациям, соответствующим следующим классическим задачам области ИБ программ.

Задача 1. Обнаружить уязвимость, которая внесена в Представлении a, если для анализа есть более позднее Представление m.

В качестве примера можно взять уже упоминаемую ранее уязвимость – пароль по умолчанию в Алгоритме программы (индекс «a» как раз и соответствует *сокр. от англ.* Algorithm, *перев. на рус.* алгоритм) внесенный туда с помощью дополнительной проверки, когда в наличии есть Машинный код этой программы (индекс «m» соответствует *сокр. от англ.* Machine, *перев. на рус.* машинный).

Запись такого Представления m с уязвимостью (внесенной операцией I^a) имеет следующий вид:

$$a < m: \langle \widehat{f^m | c^m} \rangle = T^{m-1 \to m} \ldots T^{a \to a+1} \circ I^a \circ T^{a-1 \to a} \circ \ldots \\ \circ T^{0 \to 1} \circ \langle f^0 | c^0 \rangle.$$

Попытка обнаружения указанной уязвимости с учетом того, что такая операция действует только на содержание (согласно Утверждению 12), которое меняется только при внесении уязвимостей (согласно Утверждению 5), а также следуя способу преобразования Представлений в рамках жизненного цикла (согласно Утверждению 4), будет иметь следующую запись:

$$v^a = D^a \circ \langle \widehat{f^m|c^m} \rangle =$$
$$= \left\langle (E)^a{}_f \middle| c^0 \setminus (c^0 \rhd \cdots \rhd c^a \to \hat{c}^a \rhd \cdots \rhd \hat{c}^m) \right\rangle,$$

где факт внесения уязвимости операцией I^a записан, как изменения содержания в Представлении a (согласно Утверждению 9): $c^a \to \hat{c}^a$.

Очевидно, что содержание меняется только в Представлении a и, следовательно, операция D (которая как раз и находит отличие между содержаниями) может применяться (согласно Утверждению 11) только для этого Представления:

$$v^a = D^a \circ \langle \widehat{f^m|c^m} \rangle = \left\langle (E)^a{}_f \middle| D^a \circ \hat{c}^m \right\rangle \in \emptyset,$$

где последняя часть (согласно Утверждению 11) постулирует о невозможности получения уязвимости таким образом – способом, работающем на Представлении a (в примере – для Алгоритмов), но применяемым для m (в примере – для Машинного кода).

Единственно возможным способом поиска уязвимости v^a является применение операции $(D^a)_c$ к содержанию $\hat{c}^a$, для чего в аналитической Модели существует операция обратного преобразования, применение которой для Представления m (согласно Утверждению 13) будет иметь следующий вид:

$$\hat{p}^a = \bar{T}^{a+1 \to a} \circ \ldots \circ \bar{T}^{m \to m-1} \circ \hat{p}^m.$$

В этом случае обнаружение уязвимости будет следующим:

$$v^a = D^a \circ \hat{p}^a =$$
$$\left\langle (E)^a{}_f \middle| c^0 \setminus (c^0 \rhd \cdots \rhd c^a \to \hat{c}^a \rhd \cdots \rhd \hat{c}^m \rhd \cdots \rhd \hat{c}^a) \right\rangle \notin$$
$$\notin \emptyset,$$

где $\notin \emptyset$ означает, что по данному Представлению данным способом возможно обнаружение уязвимостей.

Соответственно, класс уязвимости для примера будет следующим (согласно Утверждению 7):

$$\mathbb{C}_{v^a} = \mathbb{C}_{ML++}.$$

Описанный в задаче подход к обнаружению уязвимостей дословно звучит, как преобразование программы из Представления m к более раннему Представлению a для непосредственного поиска уязвимостей именно в нем. Такой подход является хорошо известным под общим названием «Реверс-инжиниринг программ в интересах поиска в них уязвимостей».

Далее кратко рассмотрим более сложную и реалистичную задачу.

Задача 2. Обнаружить несколько уязвимостей, которые внесены в Представления x и y, если для анализа есть более позднее Представление m.

Пример можно считать расширенной версией аналогичного для рассмотренной Задачи 1, когда имея в наличие Машинный код (Представление m), необходимо найти уязвимость Архитектуры (Представление x) с потерей функционала – отсутствие механизма шифрования (что позволяет получать несанкционированный доступ к базам данным), и Алгоритма (Представление y) с внесением функционала – вход с помощью пароля по умолчанию (что позволяет получать несанкционированный вход в систему).

Запись такого Представления имеет следующий вид:

$$x < y < m:$$
$$\hat{p}^m = \langle \widehat{f^m|c^m} \rangle = T^{m-1 \to m} \ldots T^{y \to y+1} \circ I^y \circ T^{y-1 \to y} \circ \ldots$$
$$\circ \ldots T^{x \to x+1} \circ I^x \circ T^{x-1 \to x} \circ \ldots \circ T^{0 \to 1} \circ \langle f^0|c^0 \rangle,$$

а Представления после раздельного внесения уязвимостей будут следующими:

$$\begin{cases} \hat{p}^x = I^x \circ p^x \\ \hat{p}^y = I^y \circ p^y \end{cases}.$$

Аналогичным с решением Задачи 1 образом обнаружение уязвимостей необходимо будет произвести на каждом из Представлений x и y, поскольку только на них будет работать соответствующий способ. Для этого, соответственно, потребуется обратное преобразование Представления m (в примере, Машинный код) сначала к y (в примере, к Алгоритмам), а затем к x (в примере, к Архитектуре):

$$\begin{cases} v^y = D^y \circ \hat{p}^y = \bar{T}^{y+1 \to y} \circ \ldots \circ \bar{T}^{m \to m-1} \circ \hat{p}^m \\ v^x = D^x \circ \hat{p}^x = \bar{T}^{x+1 \to x} \circ \ldots \circ \bar{T}^{y \to y-1} \circ \hat{p}^y \end{cases}.$$

Соответственно, классы уязвимости для примера будут следующими:

$$\begin{cases} \mathbb{C}_{v^x} = \mathbb{C}_{HL-+} \\ \mathbb{C}_{v^y} = \mathbb{C}_{ML++} \end{cases}.$$

Необходимо уточнить, что приведенный подход будет иметь некоторую погрешность в корректности обнаружения, поскольку уязвимости из более ранних Представлений (здесь, с индексом x) будут примешивать свое содержание к уязвимостям более поздних Представлений (здесь, с индексом y). Так, следуя примеру – в Архитектуре, полученной обратным преобразованием, помимо отсутствия механизма шифрования, будет упоминание системы аутентификации с дополнительным параметром в виде некоего пароля; а в Алгоритмах – помимо входа с помощью пароля по умолчанию будет отсутствовать целый «пласт» функционала по шифрованию данных, который потенциально должен использоваться и при хранении и проверке пароля. При этом, каждая уязвимость из другого Представления, скорее всего, не будет иметь явных признаков для обнаружения соответствующим способом.

Как результат, Представление $\hat{p}^m$ имеет сразу две уязвимости – x и y, а, следовательно, $\hat{p}^x$ и $\hat{p}^y$

после восстановления из $\hat{p}^m$, имеет такое же их количество; это в конечном итоге может привести к ошибкам в работе D^x и D^y, поскольку способы их работы ориентированы и, следовательно, наиболее результативны только для своего Представления. Для частичного нивелирования данного негативного эффекта при обнаружении уязвимостей на различных Представлениях можно применить следующий «реверс-инженерный финт» – преобразовать $\hat{p}^m$ сначала в $\hat{p}^x$, затем обнаружить уязвимость в нем, а после этого получить $\hat{p}^y$ (или прямым преобразованием из $\hat{p}^y$ или обратным их $\hat{p}^m$) для обнаружения уязвимости в нем, но уже с учетом уязвимости $\hat{p}^x$. Основная идея заключается в том, что более раннее Представление x оперирует более абстрактной базой содержания (в примере – Архитектура не использует элементы блок-схем, и, следовательно, небезопасная проверка пароля по умолчанию практически не будет влиять на обнаружение уязвимости этого Представления). Таким образом, обнаружив уязвимость в более раннем Представлении, можно минимизировать ее влияние на обнаружение уязвимостей во всех последующих Представлениях (в примере – учет того, что в Архитектуре достоверно присутствует уязвимость в виде отсутствия системы шифрования, позволит при обнаружении встроенного пароля считать, что хранение пароля не в зашифрованном виде является отдельно стоящей и уже детектированной проблемой безопасности).

Еще более реалистичной ситуацией является наличие сразу нескольких уязвимостей в одном Представлении – о чем далее.

Задача 3. Обнаружить несколько уязвимостей, которые внесены в одно Представление a, если для анализа есть более позднее Представление m.

Пример аналогичен предыдущему за исключением того, что в Алгоритмах (Представление a) присутствуют сразу две следующих уязвимости: (v_1^a) – внесение функционала в виде дополнительных проверок для входа с помощью пароля по умолчанию; (v_2^a) – изменение функционала путем изменения логина для существующей учетной записи для смены текущего администратора.

Запись такого Представления m с уязвимостью (внесенной операцией I^a) имеет следующий вид:

$$m > a: \hat{p}^m = T^{m-1\to m} \dots T^{a\to a+1} \circ I_1^a \circ I_2^a \circ T^{a-1\to a} \circ \dots \circ \circ \dots \circ T^{0\to 1} \circ p^0.$$

Применение для обнаружения D^a будет нести вероятность того, что две уязвимости выделятся, как единая – композитная, поскольку операции $I^{a1} \circ I^{a2}$ получат содержание, отличное от c^0 и трактуемое, как одна уязвимость:

$$(I_1{}^a \circ I_2^a)_c \circ c^a = c^a \to \hat{c}_2{}^a \to \hat{c}_1{}^a = c^a \to \hat{c}_{12}{}^a = = (I_{12}{}^a)_c \circ c^a,$$

где $\hat{c}_2{}^a$ и $\hat{c}_1{}^a$ – содержания, получаемые после каждого внедрения уязвимостей; $\hat{c}_{12}{}^a$ – содержание, полученное после совместного применения $I_1{}^a$ и I_2^a, которым можно сопоставить некое единое внедрение $I_{12}{}^a$ новой крупной уязвимости.

Для разрешения этой ситуации можно предположить более «умный» способ обнаружения (вместо корректного, но в данном случае тривиального нахождения отличий в содержании с помощью оператора «\»), который бы производил некоторую группировку отличий в содержании. Последнее обосновано тем, что с большой вероятностью уязвимость (в классическом понимании) является некоторым законченным и обособленным изменением в программе, а, следовательно, ее логика в базисе содержания будет строиться на близких элементах и, так или иначе, локализована в одной области. Графическая интерпретация содержания с двумя уязвимостями, поясняющая и указанный «умный» способ, в манере, аналогичной интерпретации подклассификаций уязвимостей, представлена на рисунке 2.

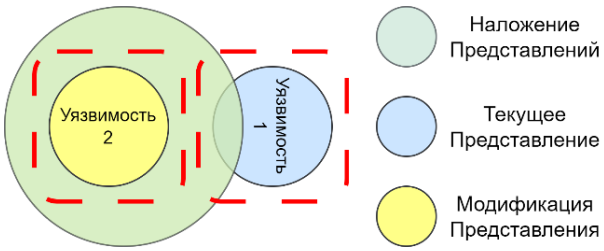


Рис. 2. Графическая интерпретация примера обнаружения двух уязвимостей в одном Представлении

Fig. 2. Graphical Interpretation of a Two Vulnerabilities Detection Example in One Representation

Согласно рисунку 2, хотя новое содержание и было получено единым изменением текущего, однако в нем все же можно выделить две группы полученных отличий – добавление (Уязвимость 1) и модификация (Уязвимость 2) элементов.

Классы же уязвимости для примера будут следующими:

$$\begin{cases} \mathbb{C}_{v_1^a} = \mathbb{C}_{ML++} \\ \mathbb{C}_{v_2^a} = \mathbb{C}_{ML\triangle\triangle} \end{cases}.$$

В заключение к приведенной аналитической Модели нужно отметить, что, хотя некоторые Утверждения (такие, как 6, 8 и 10) и примеры задач могут показаться очевидными (кстати, не без лишних оснований), однако это лишь подтверждает корректность их аналитической записи и применимость предложенного аппарата, а также обосновывает саму концепцию и эволюционные процессы в жизненном цикле программ, механизмы появления и способы обнаружения уязвимостей, а также их авторскую классификацию.

3. ЧАСТНАЯ АНАЛИТИЧЕСКАЯ МОДЕЛЬ

Сформированная аналитическая Модель жизненного цикла программы (с возможными уязвимостями) является обобщенной, поскольку предназначена для описания процессов, не зависящих от конкретных (т. е. существующих на данный момент) вариаций Представлений: их формы, содержания, способов получения и восстановления, классов уязвимостей, способов их поиска и т. п. Так, например, согласно Схеме, из Архитектуры могут быть получены Алгоритмы и Графический код, но не Дерево абстрактного синтаксиса код, а Исходный код разделяется не только на Классический, но и Метакод генерации и Сценарный. Таким образом, за обобщенной Моделью жизненного цикла, несущей более теоретический смысл, логично должно последовать создание частной Модели – которая бы отражала (т. е. моделировала) текущее состояние области программной и реверс-инженерии.

Для этого кратко запишем частную Модель на основании обобщенной Модели, указывая как формальную запись Представления, так способ и свойства его получения и восстановления, вносимые и обнаруживаемые уязвимости, а также способы их поиска (таблица 1); для сокращения записи будем, в случае необходимости, указывать в качестве номера Представления (правый верхний индекс) их перечисление. В качестве же самого номера Представления будем использовать номера на Схеме.

ТАБЛИЦА 1. Частная модель жизненного цикла программы с уязвимостями на базе эволюции ее Представлений

TABLE 1. A Life Cycle Particularly Model of the Program with Vulnerabilities Based on the Evolution of its Representations

Представления					Уязвимости		
№	Название	Формальная запись	Получение	Восстановление	Вносимые, $(\mathbb{C}_v)_{level}$	Обнаруживаемые, $(\mathbb{C}_v)_{level}$	Способ поиска, $\mathbb{C}_{detect}$
1	Идея	$p^1 \equiv \langle f^1 \mid c^1 \rangle$ $\mathbb{C}^1_{form} = \mathbb{C}_V$	Отсутствует (исходное)	$p^1 = \bar{T}^{2\to1} \circ p^2$ $\mathbb{C}^{2\to1}_{transform} = \mathbb{C}_{AT}$	Нет	Нет	Нет
2	Концептуальная модель	$p^2 \equiv \langle f^2 \mid c^2 \rangle$ $\mathbb{C}^2_{form} \in \{\mathbb{C}_V, \mathbb{C}_G\}$	$p^2 = T^{1\to2} \circ p^1$ $\mathbb{C}^{1\to2}_{transform} = \mathbb{C}_{ST}$	$p^2 = \bar{T}^{3\to2} \circ p^3$ $\mathbb{C}^{3\to2}_{transform} = \mathbb{C}_{AT}$	CL	CL	HD
3	Архитектура	$p^3 \equiv \langle f^3 \mid c^3 \rangle$ $\mathbb{C}^3_{form} \in \{\mathbb{C}_V, \mathbb{C}_G\}$	$p^3 = T^{2\to3} \circ p^2$ $\mathbb{C}^{2\to3}_{transform} = \mathbb{C}_{ST}$	$p^3 = \bar{T}^{4\to3} \circ p^4$ $p^3 = \bar{T}^{5\to3} \circ p^5$ $p^3 = \bar{T}^{6\to3} \circ p^6$ $p^3 = \bar{T}^{7\to3} \circ p^7$ $p^3 = \bar{T}^{8\to3} \circ p^8$ $\mathbb{C}^{4,5,6,7,8\to3}_{transform} = \mathbb{C}_{AT}$	HL	HL	HD
4	Двухмерная структурная схема	$p^4 \equiv \langle f^4 \mid c^4 \rangle$ $\mathbb{C}^4_{form} \in \{\mathbb{C}_P, \mathbb{C}_G\}$	$p^4 = T^{3\to4} \circ p^3$ $\mathbb{C}^{3\to4}_{transform} = \mathbb{C}_{ST}$	$p^4 = \bar{T}^{13\to4} \circ p^4$ $\mathbb{C}^{13\to4}_{transform} = \mathbb{C}_{RT}$	HL, ML	HL, ML	HD, AD
5	Функциональная диаграмма	$p^5 \equiv \langle f^5 \mid c^5 \rangle$ $\mathbb{C}^5_{form} \in \{\mathbb{C}_P, \mathbb{C}_G\}$	$p^5 = T^{3\to5} \circ p^3$ $\mathbb{C}^{3\to5}_{transform} = \mathbb{C}_{ST}$	$p^5 = \bar{T}^{13\to5} \circ p^5$ $\mathbb{C}^{13\to5}_{transform} = \mathbb{C}_{RT}$	HL, ML	HL, ML	HD, AD
6	Блок-схема	$p^6 \equiv \langle f^6 \mid c^6 \rangle$ $\mathbb{C}^6_{form} \in \{\mathbb{C}_V, \mathbb{C}_G\}$	$p^6 = T^{3\to6} \circ p^3$ $\mathbb{C}^{3\to6}_{transform} = \mathbb{C}_{ST}$	$p^6 = \bar{T}^{9\to6} \circ p^9$ $p^6 = \bar{T}^{10\to6} \circ p^{10}$ $p^6 = \bar{T}^{11\to6} \circ p^{11}$ $\mathbb{C}^{9,10,11\to6}_{transform} = \mathbb{C}_{AT}$	ML	ML	HD, AD
7	Структурограмма	$p^7 \equiv \langle f^7 \mid c^7 \rangle$ $\mathbb{C}^7_{form} \in \{\mathbb{C}_V, \mathbb{C}_G\}$	$p^7 = T^{3\to7} \circ p^3$ $\mathbb{C}^{3\to7}_{transform} = \mathbb{C}_{ST}$	$p^7 = \bar{T}^{9\to7} \circ p^9$ $p^7 = \bar{T}^{10\to7} \circ p^{10}$ $p^7 = \bar{T}^{11\to7} \circ p^{11}$ $\mathbb{C}^{9,10,11\to7}_{transform} = \mathbb{C}_{AT}$	ML	ML	HD, AD
8	Псевдокод	$p^8 \equiv \langle f^8 \mid c^8 \rangle$ $\mathbb{C}^8_{form} = \mathbb{C}_P$	$p^8 = T^{3\to8} \circ p^3$ $\mathbb{C}^{3\to8}_{transform} = \mathbb{C}_{ST}$	$p^8 = \bar{T}^{9\to8} \circ p^9$ $p^8 = \bar{T}^{10\to8} \circ p^{10}$ $p^8 = \bar{T}^{11\to8} \circ p^{11}$ $\mathbb{C}^{9,10,11\to8}_{transform} = \mathbb{C}_{AT}$	ML	ML	HD, AD
9	Классический код	$p^9 \equiv \langle f^9 \mid c^9 \rangle$ $\mathbb{C}^9_{form} = \mathbb{C}_P$	$p^9 = T^{6\to9} \circ p^6$ $p^9 = T^{7\to9} \circ p^7$ $p^9 = T^{8\to9} \circ p^8$ $p^9 = T^{10\to9} \circ p^{10}$ $\mathbb{C}^{6,7,8\to9}_{transform} = \mathbb{C}_{ST}$ $\mathbb{C}^{10\to9}_{transform} = \mathbb{C}_{GT}$	$p^9 = \bar{T}^{13\to9} \circ p^{13}$ $\mathbb{C}^{13\to9}_{transform} = \mathbb{C}_{RT}$	ML, LL	ML, LL	HD, AD
10	Метакод генерации	$p^{10} \equiv \langle f^{10} \mid c^{10} \rangle$ $\mathbb{C}^{10}_{form} = \mathbb{C}_P$	$p^{10} = T^{6\to10} \circ p^6$ $p^{10} = T^{7\to10} \circ p^7$ $p^{10} = T^{8\to10} \circ p^8$ $\mathbb{C}^{6,7,8\to10}_{transform} = \mathbb{C}_{ST}$	$p^{10} = \bar{T}^{9\to10} \circ p^9$ $\mathbb{C}^{9\to10}_{transform} = \mathbb{C}_{AT}$	ML, LL	ML, LL	HD, AD

Представления					Уязвимости		
№	Название	Формальная запись	Получение	Восстановление	Вносимые, $(\mathbb{C}_v)_{level}$	Обнаруживаемые, $(\mathbb{C}_v)_{level}$	Способ поиска, $\mathbb{C}_{detect}$
11	Сценарный код	$p^{11} \equiv \langle f^{11} \mid c^{11} \rangle$ $\mathbb{C}^{11}_{form} = \mathbb{C}_P$	$p^{11} = T^{6\to11} \circ p^6$ $p^{11} = T^{7\to11} \circ p^7$ $p^{11} = T^{8\to11} \circ p^8$ $\mathbb{C}^{6,7,8\to11}_{transform} = \mathbb{C}_{ST}$	$p^{11} = \bar{T}^{13\to11} \circ p^{13}$ $\mathbb{C}^{13\to11}_{\overline{transform}} = \mathbb{C}_{RT}$	ML, LL	ML, LL	HD, AD
12	Ассемблерный код	$p^{12} \equiv \langle f^{12} \mid c^{12} \rangle$ $\mathbb{C}^{12}_{form} = \mathbb{C}_P$	$p^{12} = T^{13\to12} \circ p^{13}$ $\mathbb{C}^{13\to12}_{transform} = \mathbb{C}_{GT}$	$p^{12} = \bar{T}^{14\to12} \circ p^{14}$ $p^{12} = \bar{T}^{15\to12} \circ p^{15}$ $\mathbb{C}^{14,15\to12}_{\overline{transform}} = \mathbb{C}_{RT}$	LL	LL	HD
13	Дерево абстрактного синтаксиса	$p^{13} \equiv \langle f^{13} \mid c^{13} \rangle$ $\mathbb{C}^{13}_{form} = \mathbb{C}_B$	$p^{13} = T^{4\to13} \circ p^4$ $p^{13} = T^{5\to13} \circ p^5$ $p^{13} = T^{9\to13} \circ p^9$ $p^{13} = T^{11\to13} \circ p^{11}$ $\mathbb{C}^{4,5,9,11\to13}_{transform} = \mathbb{C}_{GT}$	$p^{13} = \bar{T}^{12\to13} \circ p^{12}$ $p^{13} = \bar{T}^{14\to13} \circ p^{14}$ $p^{13} = \bar{T}^{15\to13} \circ p^{15}$ $\mathbb{C}^{12,14,15\to13}_{\overline{transform}} = \mathbb{C}_{AT}$	LL	LL	AD
14	Машинный код	$p^{14} \equiv \langle f^{14} \mid c^{14} \rangle$ $\mathbb{C}^{14}_{form} = \mathbb{C}_B$	$p^{14} = T^{12\to14} \circ p^{12}$ $p^{14} = T^{13\to14} \circ p^{13}$ $p^{14} = T^{15\to14} \circ p^{15}$ $\mathbb{C}^{12,13,15\to14}_{transform} = \mathbb{C}_{GT}$	Отсутствует (конечное)	AL	LL, AL	HD, AD
15	Байт-код	$p^{15} \equiv \langle f^{15} \mid c^{15} \rangle$ $\mathbb{C}^{15}_{form} = \mathbb{C}_B$	$p^{15} = T^{12\to15} \circ p^{12}$ $p^{15} = T^{13\to15} \circ p^{13}$ $\mathbb{C}^{12,13\to15}_{transform} = \mathbb{C}_{GT}$	$p^{15} = \bar{T}^{14\to15} \circ p^{14}$ $\mathbb{C}^{14\to15}_{\overline{transform}} = \mathbb{C}_{AT}$	AL	LL, AL	HD, AD

4. ЭКСПЕРИМЕНТ ПО ВОЗНИКНОВЕНИЮ УЯЗВИМОСТЕЙ

В качестве доказательства работоспособности Модели в части отражения уязвимостей приведем результаты проведенного автором ретроспективно-фактологического эксперимента. Суть эксперимента состоит в анализе баз уязвимостей на предмет отнесения моментов их возникновения к соответствующим Представлениям. Как результат, будет обосновано как существование самих Представлений, так и классов уязвимостей в них. В качестве информации об уязвимостях использовалась база CVE (*аббр. от англ.* Common Vulnerabilities and Exposures) и научные статьи из цифровой библиотеки IEEE Xplore.

Идея. Согласно Модели жизненного цикла, уязвимости в Представлении отсутствуют. В подтверждение этого соответствующих уязвимостей в базах данных найдено не было.

Концептуальная модель. Уязвимость CVE-2022-20660 заключается в хранении пароля администратора для IP-телефона Cisco (серии 7800) во флэш-памяти в открытом виде. Уязвимость относится к данному Представлению, поскольку в общей концепции этой модели телефона отсутствует понятие безопасного хранилища такого рода информации и криптографических функций в обеспечение этого. Подтверждение уязвимости: https://seclists.org/fulldisclosure/2022/Jan/34, а ее класс: $\mathbb{C}_v = \mathbb{C}_{CL-+}$.

Архитектура. Уязвимость CVE-2022-32259 заключается в том, что системные образы для установки или обновления продукта SINEMA Remote Connect Server содержат сценарии модульного тестирования, включающие в том числе конфиденциальную информацию. Уязвимость относится к данному Представлению, т. к. в архитектуре версии продукта, предназначенной для распространения и эксплуатации конечными потребителями, была оставлена подсистема тестирования. Подтверждение уязвимости: https://cert-portal.siemens.com/productcert/pdf/ssa-484086.pdf, а ее класс: $\mathbb{C}_v = \mathbb{C}_{\text{HL++}}$.

Группа «Графический код» (Двухмерная структурная схема, Функциональная диаграмма). Поскольку данное Представление, по сути, совмещает в себе два – Алгоритмы и Исходный код, то явного отнесения уязвимостей в CVE к нему найдено не было. Тем не менее, ряд частых ошибок в FBL (переходящих к уязвимостям в Представлении) был описан в статье [20]; а именно следующих: 1) неправильное использование блока таймера из-за схожести их функциональных возможностей (блок TON *аббр. от англ.* ON Delay Timer – откладывает начало события, а блок TOF *аббр. от англ.* OFF Delay Timer – откладывает остановку события); 2) ошибочный порядок входов логических операций (так, в отличие от блока AND для логической операции «И», у блока MUX для мультиплексирования их порядок имеет значение); 3) неправильное использование инвертора (часто графическое изображение операции в виде мелкого кружка приводит к ее ненужному добавлению или ошибочному опусканию на диаграмме); 4) неверная запись переменных, что ведет к некорректным присваиваниям или вычислениям (причины этого можно отнести к особенностям графической разработки алгоритмов и кода). В статье [20], помимо подтверждения указанных уязвимостей, описывается авторский метод структурного тестирования, обнаруживающий их.

Классы указанных уязвимостей (в порядке перечисления ошибок) следующие:

1) $\mathbb{C}_v = \mathbb{C}_{\mathrm{ML}\triangle?}$ (индекс «?» означает, что точный класс воздействия на информационные потоки требует уточнения, поскольку зависит от конкретной FBL-схемы);
2) $\mathbb{C}_v = \mathbb{C}_{ML\triangle\triangle}$;
3) $\mathbb{C}_v = \mathbb{C}_{ML+\triangle}$ или $\mathbb{C}_v = \mathbb{C}_{ML-\triangle}$;
4) $\mathbb{C}_v = \mathbb{C}_{\mathrm{LL}\triangle\triangle}$.

Группа «Алгоритм» (Блок-схема, Структурограмма, Псевдокод). Уязвимость CVE-2021-4021 заключается в неконтролируемом потреблении ресурсов программным продуктом для реверс-инжиниринга *Radare2* (и, как следствие, DoS-последствиям). Условием проявления уязвимости является отображение раздела файла формата ELF (64-битной версии) для процессорной архитектуры MIPS, который имеет большой раздел и заполнен нулевыми значениями. Уязвимость относится к данному Представлению, т. к. исходные алгоритмы продукта не содержали ограничений на максимально возможную анализируемую последовательность нулевых операций (так называемых NOP-инструкций), что приводило к истощению ресурсов. Подтверждение уязвимости: https://seclists.org/fulldisclosure/2022/Jan/34, а ее класс: $\mathbb{C}_v = \mathbb{C}_{ML-+}$. Трудно не отметить парадоксальность и комичность ситуации, которая заключается в том, что средство для анализа кода на предмет уязвимостей низкого уровня (в машинном коде) содержит уязвимости существенно более высокого уровня (в алгоритмах).

Группа «Исходный код» (Классический код, Метакод генерации, Сценарный код). Уязвимость CVE-2022-39288 заключается в возникновении сбоя в Web-фреймворке *fastify* при передаче некорректного Content-Type в заголовке запроса. Уязвимость относится к данному Представлению, т. к. используемый алгоритм получения парсера для определенного типа содержимого корректен, но его реализация использовала в качестве хранилища для таких парсеров JavaScript-объект, что позволяло передать туда любую строку, являющуюся ключом свойства класса Object и вызывать падение Web-сервера. Подтверждение уязвимости: https://hackerone.com/reports/1715536, а ее класс: $\mathbb{C}_v = \mathbb{C}_{LL\triangle+}$.

Ассемблерный код. Уязвимость CVE-2022-38453 заключается в потере конфиденциальности и потенциальной возможности получения полного контроля над CMS8000 (медицинское устройство для мониторинга состояния пациента https://5.imimg.com/data5/SELLER/Doc/2021/7/NM/EK/ZD/89804963/contec-cms-8000-multipara-patient-monitor-1-.pdf) из-за компиляции важных системных файлов с отладочной информацией. Уязвимость относится к данному Представлению, т. к. даже при безопасности всех более ранних Представлений компиляция программы с некорректными флагами делает генерируемый Ассемблерный код уязвимым к информационным атакам, что может привести к критическим угрозам в медицинской сфере. Подтверждение уязвимости: https://www.cisa.gov/uscert/ics/advisories/icsma-22-244-01, а ее класс: $\mathbb{C}_v = \mathbb{C}_{LL++}$.

Дерево абстрактного синтаксиса. Уязвимость CVE-2020-15294 заключается в неверной работе алгоритмов оптимизации, что в результате приводит к генерации кода, который дважды разыменовывает (т. е. получает значение по указателю) один и тот же адрес, потенциально приводя к выполнению произвольного кода. Уязвимость относится к данному Представлению, т. к. оптимизация выполняется на внутреннем Представлении входной программы в компиляторе. Подтверждение уязвимости: https://www.bitdefender.com/support/security-advisories/compiler-optimization-removal-modification-security-critical-code-vulnerability-bitdefender-hypervisor-introspection-va-9339/, а ее класс: $\mathbb{C}_v = \mathbb{C}_{LL++}$.

Машинный код. Уязвимость CVE-2021-38300 заключается в возможности выполнения произвольного кода в kernel space (*перев. на рус.* пространство ядра) на MIPS-архитектуре из-за упущения в JIT-компиляторе байт-кода cBPF (*аббр. от англ.* classic Berkeley Packet Filters – технология некоторых операционных систем, позволяющая выполнять байт-код на уровне ядра для эффективного и безопасного анализа сетевого трафика [21]), которое приводило к генерации машинного кода с недопустимым смещением при использовании условных конструкций cBPF. Уязвимость относится к данному Представлению, т. к. даже при корректности исходного кода, формирующего cBPF байт-код, среда его трансляции в машинный код (т. е. JIT-компилятор) создает угрозу ИБ. Подтверждение уязвимости: https://git.kernel.org/pub/scm/linux/kernel/git/torvalds/linux.git/commit/?id=37cb28ec7d3a36a5bace7063a3dba633ab110f8b, а ее класс: $\mathbb{C}_v = \mathbb{C}_{\mathrm{LL}\triangle+}$.

Байт-код. Уязвимость CVE-2022-31740 заключается в неверной генерации байт-кода кода утилитой компиляции WASM (*сокр. от* WebAssembly – средство для запуска скомпилированного кода в современных Web-браузерах [22]) на процессорной архитектуре Arm64, что приводило к ошибкам в распределении регистров и потенциально опасным сбоям. Уязвимость относится к данному Представлению, поскольку вносится при генерации выполняемой сборки из корректного кода высокоуровневого языка программирования (например, C++). Подтверждение уязвимости: https://bugzilla.mozilla.org/show_bug.cgi?id=1766806, а ее класс: $\mathbb{C}_v = \mathbb{C}_{AL\triangle+}$.

Исходя из приведенных и подтвержденных уязвимостей, реально существующих (или существовавших) в каждом из Представлений (или их группах), можно утверждать, что частная Модель в части уязвимостей является обоснованной.

5. ЭКСПЕРИМЕНТ ПО ЭВОЛЮЦИИ УЯЗВИМОСТЕЙ

Для демонстрации отображения уязвимостей программы в процессе эволюции ее Представлений проведем следующий практический эксперимент. В качестве программы возьмем максимально тривиальную, призванную лишь произвести указанную демонстрацию; суть программы (или другими словами, содержание) будет заключаться в делении двух чисел; естественно, операция деления должна осуществляться по всем правилам арифметики, в которой *деление на ноль неопределено*.

Искусственно заложим уязвимость в Концептуальную модель и покажем, как меняется ее отображение в процессе разработки программы (вплоть до Машинного кода) – т. е. как выглядят все Представления с и без данной уязвимости. Суть уязвимости будет заключаться в том, что пропущена проверка делителя на ноль (правила проведения арифметической операции деления нарушены). Уязвимость имеет класс $\mathbb{C}_v = \mathbb{C}_{CL-+}$, поскольку пропущен функционал, который приводит к обработке некорректных данных.

Невозможность обнаружения уязвимости в Представлениях, отличных от того, в котором она была внесена, также будет считаться подтверждением Модели и всех изложенных ранее идей данной предметной области. В качестве языка программирования выберем классический язык *C*, а в качестве Центрального процессора выполнения – Intel x86-64.

Далее кратко распишем все преобразования между Представлениями данной программы.

Идея. Замысел программы заключается в следующем: «Программа должна корректно осуществлять арифметическую операцию деления над данными пользователя».

Концептуальная модель. Форма Представления приведена на рисунке 3. Согласно Представлению, инициатор программы (Оператор) вызывает Операцию деления, которая на основании Аргументов выдает Результат вычисления. При этом, в соответствии с Идеей, деление должно выполняться согласно Правилам арифметики. Однако, в следствие внесенной уязвимости, учет правила деления на ноль был пропущен и данный элемент в итоговой модели будет отсутствовать – т. е. произошло изменение содержания в виде уязвимости класса «потеря функционала» (здесь и далее область отсутствующего функционала помечена фоном с красными наклонными линиями, которая затем заменится на области кода с рыжим фоном). Доля пропущенного элемента от всех (т. е. некий аналог «площади покрытия» уязвимостью содержания Представления посредством его формы) составляет $\frac{1}{5} = 20$ %.

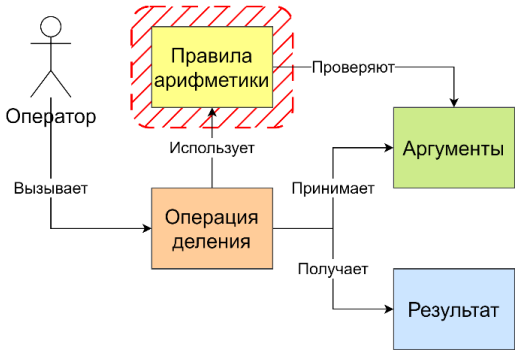


Рис. 3. Концептуальная модель программы (пример)

Fig. 3. Program Conceptual Model (Example)

Архитектура. Форма Представления приведена на рисунке 4.

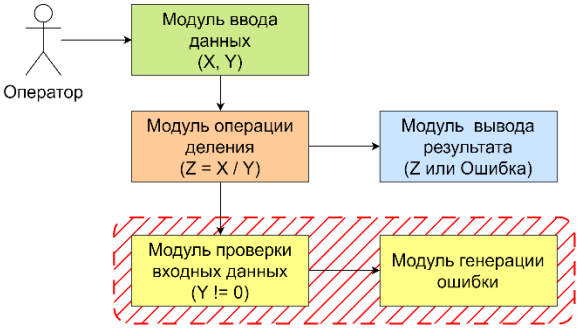


Рис. 4. Архитектура программы (примера)

Fig. 4. Program Architecture (Example)

Согласно сути Представления, полученного из Концептуальной модели, должны присутствовать точка запуска (Оператор) и 5 модулей, отвечающих как за ввод, вычисление и вывод результата деления, так и за проверку делителя на 0 с выводом ошибки. Однако, вследствие уязвимости в предыдущем Представлении, последние два модуля (отвечающие за корректность деления) были пропущены. Отметим, что несмотря на то, что уже на этом Представлении можно предположить потерю функционала, связанную с отсутствием модулей проверки и генерации ошибки, однако исходная уязвимость уже стала «распространяться» по Представлению – пропущены 2 из 6 элементов (т. е. зона охвата составила ~33 %).

Блок-схема алгоритмов. Форма Представления приведена на рисунке 5. Представление имеет вид классической Блок-схемы, по которой (после элемента Начало) производится ввод двух операндов (делимого *X* и делителя *У*), затем второй операнд сравнивается с нулевым значением. Если делитель равен 0, то выводится сообщение об ошибке; в ином случае, вычисляется частное (*Z*), которое возвращается из алгоритма. Блок-схема завершается элементом Конец. Вследствие уязвимости, в Блок-схеме отсутствует проверка делителя на 0 и вывод ошибки, а доля отсутствующего функционала составляет – $\frac{2}{7} = \sim 29$ %.

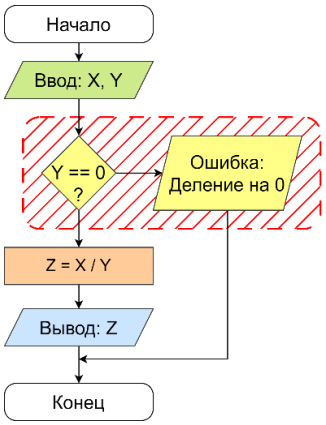


Рис. 5. Алгоритм программы (пример)

Fig. 5. Program Algorithm (Example)

Классический исходный код. В качестве данного Представления приведем одну из типовых реализаций алгоритма на языке *C*, представленную в следующем листинге:

```
#include <stdlib.h>
#include <stdio.h>
int divide (int x, int y)
{
  if (y == 0)
  {
    printf ("Divide by zero\n");
    exit (1);
  }
  int z = x / y;
  return z;
}
```

Код полностью отражает логику работы Блок-схемы программы. Область пропущенного функционала, отмеченная (здесь и далее) рыжим фоном, составляет примерно 3 из 8 строк (т. е. ~38 %), которые не включают несущественные для выполнения конструкции (в данном случае начало и завершение областей видимости переменных – посредством токенов «{» и «}»).

Дерево абстрактного синтаксиса. Представление, как правило, представляет собой совокупность графов (например, со структурой кода и потоком управления), таблиц (например, типов и символов кода) и другой служебной информации; при этом оно достаточно сильно зависит от конкретной реализации компилятора (при том, что может применяться даже для построения логической структуры тестовых документов [23]). Поэтому, без потери корректности хода эксперимента, его можно опустить.

Ассемблерный код. Представление, являющееся результатом компиляции Исходного кода (через Дерево абстрактного синтаксиса) с помощью компилятора из состава Microsoft Visual Studio Community 2019 (версия: Microsoft (R) C/C++ Optimizing Compiler Version 19.29.30145 for x86), имеет вид следующего листинга (привязка ассемблерных инструкций к Исходному коду указана в виде комментариев с префиксом «;»; конструкции ассемблера, не продуцирующие машинных инструкций, отмечены серым фоном).

```
_divide    PROC
; int divide(int x, int y)
; {
            push    ebp
            mov     ebp, esp
            push    ecx
;  if (y == 0)
;  {
            cmp     DWORD PTR _y$[ebp], 0
            jne     SHORT $LN2@divide
;    printf("Y = 0\n");
            push    OFFSET $SG10771
            call    _printf
            add     esp, 4
;    exit(1);
            push    1
            call    _exit
;  }
$LN2@divide:
;  int z = x / y;
            mov     eax, DWORD PTR _x$[ebp]
            cdq
            idiv    DWORD PTR _y$[ebp]
            mov     DWORD PTR _z$[ebp], eax
            mov     eax, DWORD PTR _z$[ebp]
$LN3@divide:
;  return z;
            mov     esp, ebp
            pop     ebp
            ret     0
; }
_divide    ENDP
```

Согласно тому, что рыжая область Ассемблерного кода, отмечающая пропущенный функционал, состоит из 7 ассемблерных строк, а весь код программы (без комментариев и не продуцирующих инструкции строк) – из 18, то область измененного содержания составляет $\frac{7}{18}$ ~39 %. Таким образом, область «распространения» уязвимости увеличилась еще больше.

Машинный код. Представление, полученное ассемблированием предыдущего, будет иметь бинарный вид, который может быть записан с помощью последовательности байт (в 16-ричной форме) следующим образом:

```
55 8B EC 51 83 7D 0C 00 75 14 68 C8 00 00 00 E8
78 00 00 00 83 C4 04 6A 01 E8 C2 00 00 00 8B 45
08 99 F7 7D 0C 89 45 FC 8B 45 FC 8B E5 5D C3
```

Очевидно, что данная форма Представления абсолютно не подходит для анализа экспертом вручную. По аналогии с предыдущими Представлениями, область с рыжим фоном соответствует уязвимости (естественно, в смысле изменения функционала); ее охват составляет $\frac{26}{47} = \sim 55$ %.

Подведем итоги эксперимента с примером данной тривиальной программы. Во-первых, базовая идея о преобразовании Представлений от Идеи до Машинного кода подтверждена. Во-вторых, уязвимость, заложенная в более раннем Представле-

нии, «живет» во всех последующих. И, в-третьих, динамика охвата уязвимостью Представления практически постоянно растет (за исключением Блок-схемы алгоритмов и отсутствующего Дерева абстрактного синтаксиса), о чем свидетельствует гистограмма на рисунке 6.

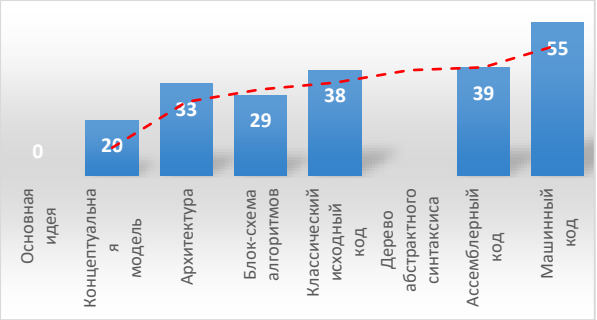


Рис. 6. Охвата уязвимостью Представлений (в процентах)

Fig. 6. Representations Coverage with Vulnerability (in Percentage)

Так, по сравнению с первым Представлением, охват уязвимости на последнем вырос в $\frac{55}{20} = \sim 2{,}75$ раза и составил 55 %. Естественно, такая динамика уязвимостей зависит от свойств конкретной программы. Однако тот факт, что даже для тривиальной программы деления двух чисел, реализуемой достаточно близкими способами, произошло сильнейшее увеличение охвата уязвимости (что в ряде случаев будет приводить к «затерянию» вредоносного кода среди безопасного, существенно затрудняя его обнаружение и «вычленение»), говорит о крайней важности учета не только статических, но и ее динамических свойств (момент возникновения и обнаружения, изменения в процессе преобразования Представлений, динамика охвата, взаимодействие с другими уязвимостями [24, 25] и т. п.).

6. ЗАКЛЮЧЕНИЕ

Проведенное в работе аналитическое моделирование программы с уязвимостями с позиции эволюции ее Представлений, основанное на результатах из предыдущей части цикла статей и подкрепленное реальными примерами, позволяет более полно взглянуть на динамику процесса создания программ и существования уязвимостей с формальной точки зрения, что приводит к ряду следующих умозаключений.

Во-первых, все объекты и процессы, упоминаемые при описании Схемы жизненного цикла, а также производные от них, переведены в разряд Утверждений и записаны в аналитическом виде. Как результат, была построена обобщенная аналитическая Модель, находящаяся в строгом соответствии с предыдущими выкладками.

Во-вторых, учет способов преобразования между Представлениями при создании программ и их восстановлении позволил создать частную аналитическую Модель, отражающую текущее состояние области программной инженерии.

В-третьих, введение в Модели понятия уязвимости и операций, связанных с ними (внесение, обнаружение, классификация) расширил понимание этого достаточно сложного и противоречивого объекта области ИБ программ.

Проведенные эксперименты по возникновению уязвимостей и их эволюции между Представлениями обосновывают сделанные выводы. Продолжение же работы должно быть направлено на теоретико-практический аспект способов обратного преобразования Представлений, что (как уже было многократно сказано) существенно повысит возможности по обнаружению и нейтрализации уязвимостей в программах. В интересах этого автор видит перспективным развитие отдельного направления *интеллектуального реверс-инжиниринга*, например, на базе генетической декомпиляции [26–28]. Основная идея последней заключается в итеративном приближении предыдущих Представлений к таким форме и содержанию, чтобы их прямое преобразование давало бы в точности заданное Представление. Как результат, удастся снизить необходимость в сверхвостребованных экспертах по безопасности программного кода.

Список источников

1. Kotenko I., Izrailov K., Buinevich M. Static Analysis of Information Systems for IoT Cyber Security: A Survey of Machine Learning Approaches // Sensors. 2022. Vol. 22. Iss. 4. P. 1335. DOI:10.3390/s22041335

2. Trevizan R.D., Obert J., De Angelis V., Nguyen Tu.A., Rao V.S., Chalamala B.R. Cyberphysical Security of Grid Battery Energy Storage Systems // IEEE Access. 2022. Vol. 10. PP. 59675–59722. DOI:10.1109/ACCESS.2022.3178987

3. Cho C.-S., Chung W.-H., Kuo S.-Y. Cyberphysical Security and Dependability Analysis of Digital Control Systems in Nuclear Power Plants // IEEE Transactions on Systems, Man, and Cybernetics: Systems. 2016. Vol. 46. Iss. 3. PP. 356–369. DOI:10.1109/TSMC.2015.2452897

4. Израилов К.Е. Моделирование программы с уязвимостями с позиции эволюции ее представлений. Часть 1. Схема жизненного цикла // Труды учебных заведений связи. 2023. Т. 9. № 1. С. 75–93. DOI:10.31854/1813-324X-2023-9-1-75-93

5. Монастырная В.С., Фролов В.В. Визуальный язык дракон и его применение // Актуальные проблемы авиации и космонавтики. 2016. Т. 2. № 12. С. 78–79.

6. Пароджанов В.Д. Алгоритмические языки и программирование: ДРАКОН: учебное пособие для среднего профессионального образования. М.: Издательство Юрайт, 2023. 436 с.

7. Долидзе А.Н. Обзор специфических функций языка FBD на примере программируемых реле Logo! // Инженерный вестник Дона. 2022. № 11(95). С. 1–10.

8. Pardo M.X.C., Ferreiro G.R. SFC++: A Tool for Developing Distributed Real-Time Control Software // Microprocessors and Microsystems. 1999. Vol. 23. Iss. 2. PP. 75–84. DOI:10.1016/S0141-9331(99)00015-0

9. Ахмерова А.Н. Языки программирования контроллеров. Особенности применения языков FBD, LD // Научный аспект. 2019. Т. 3. № 3. С. 340–345.

10. Nassi I., Shneiderman B. Flowchart techniques for structured programming // SIGPLAN Notices. Vol. 8. Iss. 8. PP. 12–26. DOI:10.1145/953349.953350

11. Басов А.С. Классификация языков программирования и их особенности // Вестник науки. 2020. Т. 2. № 8(29). С. 95–101.

12. Морозов Д.П., Слепнев А.В. Разработка анализатора кода C, C++ на языке Python с использованием Lex, Yacc // 74-я региональная научно-техническая конференция студентов, аспирантов и молодых ученых. Студенческая весна – 2020 (Санкт-Петербург, Россия, 26–27 мая 2020). СПб.: СПбГУТ, 2020. С. 28–32.

13. Lee W.I., Lee G. From natural language to Shell Script: A case-based reasoning system for automatic UNIX programming // Expert Systems with Applications. 1995. Vol. 9. Iss. 1. PP. 71–79. DOI:10.1016/0957-4174(94)00050-6

14. Пирогов В. Ассемблер для Windows. Санкт-Петербург: БХВ-Петербург, 2012. 896 с.

15. Капустин Д.А., Швыров В.В., Шулика Т.И. Статический анализ корпуса исходных кодов Python-приложений // Программная инженерия. 2022. Т. 13. № 8. С. 394–403. DOI:10.17587/prin.13.394-403

16. Кричанов М.Ю., Чепцов В.Ю. Защищенная UEFI-прошивка для виртуальных машин // Системный администратор. 2021. № 11(228). С. 75–81.

17. Макаров А.В., Скоробогатов С.Ю., Чеповский А.М. Common Intermediate Language и системное программирование в Microsoft.NET: учебное пособие. Москва, Саратов: Интернет-Университет Информационных Технологий (ИНТУИТ), Ай Пи Ар Медиа, 2020. 397 с.

18. Красов А.В., Шариков П.И. Методика защиты байт-кода Java-программы от декомпиляции и хищения исходного кода злоумышленником // Вестник Санкт-Петербургского государственного университета технологии и дизайна. Серия 1: Естественные и технические науки. 2017. № 1. С. 47–50.

19. Израилов К.Е., Татарникова И.М. Подход к анализу безопасности программного кода с позиции его формы и содержания // VIII Международная научно-техническая и научно-методическая конференция «Актуальные проблемы инфотелекоммуникаций в науке и образовании» (Санкт-Петербург, Россия, 27–28 февраля 2019). СПб: СПбГУТ, 2019. С. 462–467.

20. Eunkyoung J., Seungjae J., Hojung B., Sungdeok C., Junbeom Y., Geeyong P., et al. Testing of Timer Function Blocks in FBD // Proceedings of 13th Asia Pacific Software Engineering Conference (APSEC'06, Bangalore, India, 06–08 December 2006). IEEE, 2006. PP. 243–250. DOI:10.1109/APSEC.2006.55

21. McCanne S., Jacobson V. The BSD Packet Filter: A New Architecture for User-Level Packet Capture // Proceedings of the Winter USENIX Technical Conference, San Diego, USA, 25–29 January 1993. USENIX Association, 1993.

22. Kim M., Jang H., Shin Y. Avengers, Assemble! Survey of WebAssembly Security Solutions // Proceedings of 15th International Conference on Cloud Computing (CLOUD, Barcelona, Spain, 10–16 July 2022). IEEE, 2022. PP. 543–553. DOI:10.1109/CLOUD55607.2022.00077

23. Чувилин К.В. Параметрический подход к построению синтаксических деревьев для частично формализованных текстовых документов // Машинное обучение и анализ данных. 2016. Т. 2. № 2. С. 201–217.

24. Буйневич М.В., Израилов К.Е. Антропоморфический подход к описанию взаимодействия уязвимостей в программном коде. Часть 1. Типы взаимодействий // Защита информации. Инсайд. 2019. № 5(89). С. 78–85.

25. Буйневич М.В., Израилов К.Е. Антропоморфический подход к описанию взаимодействия уязвимостей в программном коде. Часть 2. Метрика уязвимостей // Защита информации. Инсайд. 2019. № 6(90). С. 61–65.

26. Израилов К.Е. Концепция генетической декомпиляции машинного кода телекоммуникационных устройств // Труды учебных заведений связи. 2021. Т. 7. № 4. С. 10–17. DOI:10.31854/1813-324X-2021-7-4-95-109

27. Израилов К.Е. Применение генетических алгоритмов для декомпиляции машинного кода // Защита информации. Инсайд. 2020. № 3(93). С. 24–30.

28. Израилов К.Е., Романов Н.Е. Применение генетического алгоритма для реверс-инжиниринга машинного кода // XI Международная научно-техническая и научно-методическая конференция «Актуальные проблемы инфотелекоммуникаций в науке и образовании» (Санкт-Петербург, Россия, 15–16 февраля 2022). СПб: СПбГУТ, 2022. С. 239–243.

References

1. Kotenko I., Izrailov K., Buinevich M. Static Analysis of Information Systems for IoT Cyber Security: A Survey of Machine Learning Approaches. *Sensors*. 2022;22(4):1335. DOI:10.3390/s22041335 (in Russ.)

2. Trevizan R.D., Obert J., De Angelis V., Nguyen Tu.A., Rao V.S., Chalamala B.R. Cyberphysical Security of Grid Battery Energy Storage Systems. *IEEE Access*. 2022;(10):59675–59722. DOI:10.1109/ACCESS.2022.3178987

3. Cho C.-S., Chung W.-H., Kuo S.-Y. Cyberphysical Security and Dependability Analysis of Digital Control Systems in Nuclear Power Plants. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*. 2016;46(3):356–369. DOI:10.1109/TSMC.2015.2452897

4. Izrailov K. Modeling a Program with Vulnerabilities in the Terms of Its Representations Evolution. Part 1. Life Cycle Scheme. *Proc. of Telecom. Universities*. 2023;9(1):75–93. (In Russ.) DOI:10.31854/1813-324X-2023-9-1-75-93

5. Monastyrnaya V.S., Frolov V.V. Visual language dragon and it is application. *Aktual'nye problemy aviacii i kosmonavtiki*. 2016;2(12):78–79. (in Russ.)

6. Parondzhanov V.D. *Algorithmic Languages and Programming: DRAGON*. Moscow: Yurajt Publ.; 2023. 436 p. (in Russ.)

7. Dolidze A.N. Overview of specific functions of the FBD language using the example of Logo! *Engineering journal of Don*. 2022;11(95):1–10. (in Russ.)

8. Pardo M.X.C., Ferreiro G.R. SFC++: A Tool for Developing Distributed Real-Time Control Software. *Microprocessors and Microsystems*. 1999;23(2):75–84. DOI:10.1016/S0141-9331(99)00015-0

9. Akhmerova A.N. Controller programming languages. Features of the application of the languages. *Nauchnyj aspekt.* 2019;3(3):340–345. (in Russ.)

10. Nassi I., Shneiderman B. Flowchart techniques for structured programming. *SIGPLAN Notices*. 8(8):12–26. DOI:10.1145/953349.953350

11. Basov A.S. Classification of Programming Languages and their Features. *Vestnik nauki*. 2020;2(8):95–101. (in Russ.)

12. Morozov D.P., Slepnev A.V. Development of C, C++ code analyzer in Python using Lex, Yacc. *Proceedings of the 74th Regional Scientific and Technical Conference of Students, Graduate Students and Young Scientists "Student Spring – 2020", 26–27 May 2020, St. Petersburg, Russia.* St. Petersburg: The Bonch-Bruevich Saint Petersburg State University of Telecommunications Publ.; 2020. p.28–32. (in Russ.)

13. Lee W.I., Lee G. From natural language to Shell Script: A case-based reasoning system for automatic UNIX programming. *Expert Systems with Applications*. 1995;9(1):71–79. DOI:10.1016/0957-4174(94)00050-6

14. Pirogov V. *Assembler for Windows*. BHV-Petersburg Publ.; 2012. 896 p. (in Russ.)

15. Kapustin D.A., Shvyrov V.V., Shulika T.I. Static analysis of the source code of python applications. *Software Engineering*. 2022;13(8):394–403. (in Russ.) DOI:10.17587/prin.13.394-403

16. Krichanov M.Y., Cheptsov V.Y. Secure UEFI firmware for virtual machines. *Sistemnyj administrator*. 2021;11(228): 75–81. (in Russ.)

17. Makarov A.V., Skorobogatov S.Y., Chepovskii A.M. *Common Intermediate Language and system programming in Microsoft.NET*. Moscow, Saratov: Internet University of Information Technologies Publ.; Ai Pi Ar Media Publ.; 2020. 397 p. (in Russ.)

18. Krasov A.V., Sharikov P.I. Methods of protection byte code java-programs from decompilation and theft of source code by an attacker. *Vestnik of St. Petersburg State University of Technology and Design. Series 1: Natural and technical Sciences.* 2017;(1):47–50. (in Russ.)

19. Izrailov K., Tatarnikova I. An Approach to Analyzing the Security of a Software Code from the Standpoint of Its Form and Content. *Proceedings of the VIIth International Conference on Infotelecommunications in Science and Education, 27–28 February 2019, St. Petersburg, Russia*. St. Petersburg: The Bonch-Bruevich Saint-Petersburg State University of Telecommunications Publ.; 2019. p.462–467. (in Russ.)

20. Eunkyoung J., Seungjae J., Hojung B., Sungdeok C., Junbeom Y., Geeyong P., et al. Testing of Timer Function Blocks in FBD. *Proceedings of 13th Asia Pacific Software Engineering Conference, APSEC'06, 06–08 December 2006, Bangalore, India*. IEEE; 2006. p.243–250. DOI:10.1109/APSEC.2006.55

21. McCanne S., Jacobson V. The BSD Packet Filter: A New Architecture for User-Level Packet Capture. *Proceedings of the Winter USENIX Technical Conference, 25–29 January 1993, San Diego, USA*. USENIX Association; 1993.

22. Kim M., Jang H., Shin Y. Avengers, Assemble! Survey of WebAssembly Security Solutions. *Proceedings of 15th International Conference on Cloud Computing, CLOUD, 10–16 July 2022, Barcelona, Spain*. IEEE; 2022. p.543–553. DOI:10.1109/CLOUD55607.2022.00077

23. Chuvilin K.V. Parametric Approach to the Construction of Syntax Trees for Partially Formalized Text Documents. *Machine Learning and Data Analysis*. 2016;2(2):201–217. (in Russ.)

24. Buinevich M.V., Izrailov K.E. Anthropomorphic approach to describing the interaction of vulnerabilities in program code. Part 1. Types of interactions. *Zaŝita informacii. Inside*. 2019;5(89):78–85. (in Russ.)

25. Buinevich M.V., Izrailov K.E. Anthropomorphic approach to describing the interaction of vulnerabilities in program code. Part 2. Vulnerability metric. *Zaŝita informacii. Inside*. 2019;6(90):61–65. (in Russ.)

26. Izrailov K. The Genetic Decompilation Concept of the Telecommunication Devices Machine Code. *Proc. of Telecom. Universities*. 2021;7(4):10–17. DOI:10.31854/1813-324X-2021-7-4-95-109 (in Russ.)

27. Izrailov K.E. Applying of genetic algorithms to decompile machine code. *Zaŝita informacii. Inside*. 2020;3(93):24–30. (in Russ.)

28. Izrailov K.E., Romanov N.E. Application of genetic algorithm for reverse engineering of machine code. *Proceedings of the XIth International Conference on Infotelecommunications in Science and Education, 15–16 February 2022, St. Petersburg, Russia*. St. Petersburg: The Bonch-Bruevich Saint-Petersburg State University of Telecommunications Publ.; 2022. p. 239–243. (in Russ.)

Статья поступила в редакцию 20.02.2023; одобрена после рецензирования 27.02.2023; принята к публикации 25.03.2023.
The article was submitted 20.02.2023; approved after reviewing 27.02.2023; accepted for publication 25.03.2023.

Информация об авторе:

ИЗРАИЛОВ Константин Евгеньевич

кандидат технических наук, доцент, старший научный сотрудник Санкт-Петербургского Федерального исследовательского центра Российской академии наук
https://orcid.org/0000-0002-9412-5693

Научная статья
УДК004:519.854
DOI:10.31854/1813-324X-2023-9-2-112-127

Иерархическая модель и алгоритм оптимизации решений при распределенном хранении и обработке данных

Кирилл Викторович Кротов, krotov_k1@mail.ru

Севастопольский государственный университет,
Севастополь, 299053, Российская Федерация

Аннотация: *Задача оптимизации распределенного хранения и обработки данных является трудноразрешимой за ограниченное время. В связи с этим для ее решения применен иерархический подход, предусматривающий представление обобщенной задачи в виде совокупности иерархически упорядоченных подзадач, для каждой из которых на соответствующем ей уровне иерархии определяются локально оптимальные решения. Для оптимизации решений по распределенному хранению и обработке данных сформирована модель процесса в виде совокупности иерархически упорядоченных компонент, а также математическая модель иерархической игры, представляющая собой способ оптимизации решений на уровнях иерархии. С целью определения эффективных решений на уровнях иерархии разработан алгоритм локальной оптимизации решений, в основу которого положены генетические алгоритмы. Построение расписаний обработки данных, назначенных на вычислительные устройства, реализуется с использованием предложенной эвристической процедуры. Применение разработанных моделей процесса распределенного хранения и обработки данных, модели иерархический игры и алгоритмов оптимизации решений позволили значительно увеличить размерность задачи, учесть при оптимизации решений на уровнях иерархии параметры, характеризующие каналы передачи данных, минимизировать количество неиспользованных ресурсов.*

Ключевые слова: *иерархическая игра, ограничения на объемы распределенных устройств хранения, оптимизация решений по распределенному хранению и распределенной обработке данных, генетические алгоритмы.*

Ссылка для цитирования: Кротов К.В. Иерархическая модель и алгоритм оптимизации решений при распределенном хранении и обработке данных // Труды учебных заведений связи. 2023. Т. 9. № 2. С. 112–127. DOI:10.31854/1813-324X-2023-9-2-112-127

Hierarchical Model and Decision Optimization Algorithm for Distributed Data Storage and Processing

Kirill Krotov, krotov_k1@mail.ru

[1]Sevastopol State University
Sevastopol, 299053, Russian Federation

Abstract: *The task of optimizing distributed data storage and processing is difficult to solve in a limited time. In this regard, a hierarchical approach has been applied to solve it, which provides for the presentation of a generalized problem in the form of a set of hierarchically ordered subtasks, for each of which locally optimal solutions are determined at the appropriate hierarchy level. To optimize solutions for distributed data storage and processing, a process model has been formed, presented in the form of a set of hierarchically ordered components, a mathematical model of a hierarchical game, which is a way to optimize solutions at hierarchy levels. In order to determine effective solutions at hierarchy levels, an algorithm for local optimization of solutions based on genetic algorithms*

has been developed. The construction of data processing schedules assigned to computing devices is implemented using the proposed heuristic procedure. The application of the developed models of the distributed data storage and processing process, hierarchical game models and algorithms for optimizing solutions made it possible to significantly increase the dimension of the problem, take into account the parameters characterizing data transmission channels when optimizing solutions at hierarchy levels, and minimize the amount of unused resources.

Keywords: *hierarchical game, restrictions on the volume of distributed storage devices, optimization of solutions for distributed storage and distributed data processing, genetic algorithms*

For citation: Krotov K. Hierarchical Model and Decision Optimization Algorithm for Distributed Data Storage and Processing. *Proc. of Telecom. Universities.* 2023;9(2):112–127. (in Russ.) DOI:10.31854/1813-324X-2023-9-2-112-127

Введение

В настоящее время активно развивается подход к обработке данных, предполагающий, что устройства их хранения и выполнения вычислений с ними являются географически распределенными. Указанный подход к обработке данных предусматривает распределенное их хранение и распределенную их обработку, в смысле – выполнение вычислений (РХ-РО). Предполагается, что для решения некоторого пула задач обработки данных выделяются ресурсы ограниченного объема как для их хранения, так и для вычислений. Также следует учитывать, что хранение данных на устройствах и их обработка связаны с взиманием платы за оказываемые услуги. Отсюда рассматриваемая постановка задачи типа «РХ-РО» характеризуется особенностями, связанными с наличием ограничений на ресурсы, а также с назначением стоимости за оказание соответствующих услуг.

В то же время немаловажным фактором является эффективное использование ресурсов, реализующих передачу данных между устройствами хранения и обработки, которое также характеризуется стоимостными показателями. Таким образом, реализация РХ-РО обеспечивается использованием трех видов ресурсов. В связи с необходимостью эффективного применения ограниченных ресурсов, а также с целью минимизации финансовых затрат требуется оптимизировать их (данных) распределение по устройствам хранения и назначение вычислительных устройств (ВУ) для их обработки.

Современное состояние научных исследований, направленных на планирование распределенного выполнения заданий (обработки данных на распределенных ВУ) связано в основном с применением различных эвристических подходов. В [1] рассматривается применение известных эвристик для выбора задания, назначаемого на устройство в GRID-системах: «задание с наименьшим временем обработки – первым» (SPTF, *аббр. от англ.* Shortest-Processing-Time-First), «короткое задание – первым» (SJF, *аббр. от англ.* Short-Job-First), «задание с самым длинным временем обработки – первым» (LPTF, *аббр. от англ.* Longest-Processing-Time-First), «задание самого большого размера – первым» (LSF, *аббр. от англ.* Largest-Size-First), «задание с самым ранним директивным сроком окончания выполнения – первым» (EDF, *аббр. от англ.* Earliest-Deadline-First). Также в [1] предложен способ упорядочивания заданий в очереди на выполнение с использованием специальных метрик, выражения для вычисления которых предусматривают использование параметров: директивного срока, длительности обработки данных, значения текущего времени. Кроме того, в [1] рассматривается применение эвристик для выбора устройства, на которое может быть назначено задание из головы очереди: эвристика «выбор наиболее подходящего» (*от англ.* Best-Fit) выбирает узел, который имеет наименьшее количество доступных ресурсов и может выполнить обработку рассматриваемых данных; эвристика «выбор первого подходящего» (*от англ.* First-Fit) планирует задания по первому подходящему ресурсу из списка доступных; эвристика «самый быстрый ресурс – первый» (FRF, *аббр. от англ.* Fastest-Resource-First) выбирает самый быстрый ресурс из списка доступных и назначает его для выполнения задания; эвристика «минимально загруженный ресурс – первый» (MLF, *аббр. от англ.* Min-Loaded-First) выбирает ресурс, имеющий максимальную неиспользуемую мощность процессора.

Развитию эвристического подхода к планированию в GRID-системах посвящена работа [2], в которой рассматриваются эвристики: 1) UDA (*аббр. от англ.* User-Defined-Attributes, определяемый пользователем атрибут) – назначает устройству задачу с наилучшим ожидаемым временем ее выполнения, независимо от его (устройства) доступности; 2) алгоритм Min-Min–задача с самым минимальным временем выполнения назначается на соответствующий ресурс; 3) алгоритм Max-Min–задача с максимальным (среди минимальных) временем выполнения назначается на соответствующий ресурс.

В работе [3] также рассматривается применение эвристик Min-Min и Max-Min к планированию в GRID-системах; анализируются эвристические правила:

1) «минимальное время выполнения – первым» (MET, *аббр. от англ.* Minimum-Execution-Time), ко-

торое назначает задание устройству, имеющему наименьшее время выполнения для этого задания;

2) «самое длинное задание – на самый быстрый ресурс» (LJFR, *аббр. от англ.* Longest-Job-to-Fastest-Resource);

3) «самое короткое задание – на самый быстрый ресурс» (SJFR, *аббр. от англ.* Shortest-Job-to-Fastest-Resource);

4) «минимальная относительная стоимость выполнения задания – первой» (RC, *аббр. от англ.* Relative-Cost) – предполагает вычисление значений параметров статической и динамической относительной стоимости и упорядочивание заданий в соответствии с этими значениями.

Исследованию рассмотренных выше эвристик (в том числе MET, MLF, FRF, Min-Min и Max-Min) применительно к планированию в GRID-системах посвящена работа [4]. Также упоминается о возможности применения метаэвристических алгоритмов (генетические алгоритмы, муравьиные колонии, имитация отжига) к решению задач планирования в рассмотренных системах. Однако непосредственная адаптация этих алгоритмов для решения указанных задач планирования в этой работе не рассматривается.

В работе [5] рассматриваются различные политики (эвристические правила) назначения ресурсов заданиям в облачных средах. Это политика «первого доступного кэша» (FCA, *аббр. от англ.* First-Cache-Available), политика «максимального количества попаданий кэша» (MCH, *аббр. от англ.* Max-Cache-Hit), «максимальное количество вычислений» (MCU, *аббр. от англ.* Max-Compute-Util) и «хорошее вычисление кэша» (GCC, *аббр. от англ.* Good-Cache-Computing). Политика FCA игнорирует информацию о местоположении данных при выборе исполнителя для задачи; она обеспечивает выбор первого доступного исполнителя (устройства) и не предоставляет исполнителю никакой информации о расположении данных, необходимых для задачи; исполнитель должен получать все эти данные из постоянного хранилища при каждом доступе. Политика MCH использует информацию о расположении данных для отправки каждой задачи исполнителю с наибольшим объемом данных, необходимых для ее решения. Если исполнитель занят, то отправка откладывается до тех пор, пока он не станет доступным. Политика MCU использует информацию о местоположении данных, пытаясь максимизировать использование ресурсов даже при потенциально более высоких затратах на перемещение данных. Она назначает задачу доступному исполнителю, предпочитая исполнителей, находящихся «ближе» к необходимым для ее решения данным. Политика GCC – гибридная политика по отношению к MCH и MCU; она устанавливает порог минимального использования процессора с целью определения необходимости использования одной из указанных политик (MCH или MCU).

В работе [6] наряду с уже рассмотренными эвристиками используются следующие правила для планирования выполнения заданий в вычислительных кластерах: «наименее подходящий» (WF, *аббр. от англ.* Worst-Fit); случайное назначение (RF, *аббр. от англ.* Random-Fit), а также методы предварительного резервирования и обратного заполнения. Метод предварительного резервирования использует информацию о времени выполнения, предоставленную пользователем, чтобы зарезервировать ресурсы процессора и памяти, и соответственно сгенерировать расписание. Метод обратного заполнения является улучшением алгоритма планирования с полным распределением задач по ресурсам.

В работе [7] рассматривается способ определения количества задач, входящих в задания (работы), назначаемые для выполнения на разных ВУ. При определении указанного количества задач учитывается степень параллелизма (то есть количество параллельно выполняющихся задач, входящих в задание/работу), а также производительность (скорость выполнения вычислений) как отдельных процессоров, так и производительность всей системы. Определенные таким образом совокупности задач назначаются соответствующим ВУ для выполнения.

Рассмотренный в [8] алгоритм обеспечивает назначение виртуальных машин для выполнения заданий с учетом их (заданий) бюджета. Алгоритм предполагает первоначально формирование списка виртуальных машин, которые могут быть назначены для выполнения заданий. Для каждой виртуальной машины из этого списка с учетом характеристик заданий (количество операций в программах, размеры входного и выходного файлов), а также характеристик самих виртуальных машин (пропускные способности каналов, связывающих устройства, объемы оперативной и постоянной памяти, выделяемых для хранения данных, стоимости передачи данных по каналам и хранения) определяется стоимость их (заданий) выполнения на соответствующих устройствах. В том случае, если стоимость выполнения задания на виртуальной машине не превышает выделенного для него бюджета, то рассматриваемая машина назначается для реализации вычислений с этим заданием. Здесь стоимость выполнения заданий на виртуальных машинах (и формула, позволяющая вычислить эту стоимость) является своего рода эвристикой, используемой при планировании.

В [9] предложен алгоритм назначения ВУ в облачной среде с учетом заданных директивных сроков окончания выполнения заданий и ограничений на бюджеты, связанные с их выполнением. Задан

граф следования заданий при реализации вычислительного процесса (граф передачи управления между заданиями при реализации вычислительного процесса). Алгоритм предусматривает первоначальное распределение заданий по уровням независимого (параллельного) выполнения – формирование соответствующей формы, в соответствии с которой реализуется назначение заданий каждого уровня (при движении сверху-вниз по полученной форме) на ВУ с учетом бюджета выполнения и с последующим определением сроков их завершения. При распределении заданий по устройствам учитываются только временные и стоимостные параметры их выполнения. Рассмотренный алгоритм не предусматривает оптимизации решений и может быть отнесен к эвристическим.

В целом выполненный анализ существующих методов планирования распределенного выполнения заданий на ВУ позволил определить следующие их недостатки:

– рассмотренные алгоритмы и методы реализуют планирование выполнения заданий на устройствах с использованием эвристического подхода (эвристических правил и алгоритмов), то есть не предусматривают поиска приближенно оптимальных или локально оптимальных решений;

– при распределении заданий на устройства большинство алгоритмов учитывают только один вид ресурсов – процессорное время, но не учитывают ресурсы других видов – каналы передачи данных и хранилища данных (в том числе ограниченного объема).

В связи с этим разработка новых методов и алгоритмов планирования распределенного хранения и распределенной обработки данных при учете передачи данных между устройствами и учете стоимостных характеристик реализации указанных операций является актуальной.

Задача оптимизации решений по размещению данных на устройствах хранения ограниченного объема, по назначению ВУ для их обработки (с учетом использования каналов передача между указанными устройствами) является NP-трудной [1]. В связи с этим необходима разработка приближенных методов оптимизации решений или методов локальной оптимизации решений. Одним из возможных способов решения указанной задачи оптимизации является иерархический подход [10, 11], который предусматривает выделение в обобщенной задаче оптимизации совокупности подзадач, каждая из которых решается на назначенном ей уровне иерархии; реализация указанного подхода предполагает:

– определенный порядок формирования решений – первоначально решение формируется на вышестоящем уровне, затем на нижестоящем уровне выбирается лучшее (в частном случае, локально-оптимальное);

– необходимость обмена решениями между уровнями – решение, сформированное на вышестоящем уровне, передается на нижестоящий для определения локально-оптимального решения, которое передается на вышестоящий для вычисления на его основе оценки оптимальности;

– зависимость оценки оптимальности решений на вышестоящем уровне от решения на нижестоящем – обобщенное оптимальное решение задачи формируется путем определения оптимальных решений на каждом из уровней иерархии.

В соответствии с указанными особенностями иерархического подхода для решения сформулированной задачи может быть применен аппарат теории иерархических игр. В связи с этим получение эффективных решений по размещению данных в хранилищах, по назначению ВУ для их обработки обеспечивается разработкой:

– математической модели процесса распределенных хранения и обработки данных (при ограничениях на размеры хранилищ и установлении стоимостных характеристик по оказанию услуг, связанных с хранением, обработкой и передачей данных);

– математической модели иерархической игры оптимизации решений по распределенному хранению и распределенной обработке данных как способа определения эффективных (в частном случае, локально-оптимальных) решений;

– метода (алгоритма) поиска локально-оптимальных решений на каждом из уровней иерархии;

– алгоритма построения расписаний обработки данных, назначенных на соответствующие ВУ.

Таким образом, основное предназначение выполняемого исследования состоит в значительном увеличении размерности рассматриваемых задач, а также в гарантированном получении эффективных решений при различных значениях их (задач) входных данных.

1. Математическое моделирование процессов и синтез математической модели оптимизации решений по распределенному хранению и обработке данных

Для синтеза математической модели процесса распределенного хранения и обработки данных в рассмотрение введены обозначения для входных данных и параметров задачи:

1) i – идентификатор типа данных, распределенное хранение и обработка которых выполняется в системе ($i = \overline{1, N}$);

2) d_i – объем данных i-го типа, которые должны быть размещены в устройствах хранения ($i = \overline{1, N}$);

3) m – идентификатор устройства распределенного хранения данных ($m = \overline{1, M}$);

4) v_m – размер m-го хранилища данных ($m = \overline{1, M}$), являющийся задаваемым при решении задачи оптимизации; $V = (v_m | \ m = \overline{1, M})$– вектор-строка размеров m-х хранилищ данных;

5) s_m – стоимость хранения единицы данных в единицу времени на m-м устройстве хранения; $S = (s_m | \ m = \overline{1, M})^T$ – вектор-столбец стоимостей хранения данных разных типов в единицу времени на m-х устройствах;

6) sh_m – размер штрафа за единицу неиспользованного ресурса на m-ом устройстве хранения; $Sh = (sh_m | \ m = \overline{1, M})$– вектор-строка штрафов за единицу неиспользованного ресурса на m-х устройствах;

7) l – идентификатор устройства распределенной обработки данных ($l = \overline{1, L}$);

8) t_{il}– длительность обработки данных i-го типа на l-м ВУ ($i = \overline{1, N}; l = \overline{1, L}$); $T = \|t_{il}\|_{N \times L}$– матрица длительностей обработки данных i-х типов на l-х ВУ;

9) w_l – стоимость единицы времени обработки данных на l-м ВУ ($l = \overline{1, L}$);

10) c_{ml} – пропускная способность канала передачи данных между m-м устройством хранения и l-м устройством обработки (в том случае, если между m-м устройством хранения и l-м устройством обработки отсутствует канал передачи данных, то $c_{ml} = 0$); $C = \|c_{ml}\|_{M \times L}$ – матрица пропускных способностей каналов передачи данных между m-ми устройствами хранения и l-ми устройствами обработки;

11) b_{ml} – длина канала передачи данных между m-м устройством хранения и l-м устройством обработки ($b_{ml} = 0$, если канал передачи данных отсутствует); $B = \|b_{ml}\|_{M \times L}$ – матрица длин каналов передачи данных между m-ми устройствами хранения и l-ми устройствами обработки;

12) $b_{\max}$ – максимальная длина канала передачи данных, используемая для вычисления значений коэффициентов длин каналов θ_{ml} ($m = \overline{1, M}$; $l = \overline{1, L}$);

13) θ_{ml} – коэффициент длины канала передачи данных между m-м устройством хранения и l-м устройством обработки (вычисляется в соответствии с формулой: $\theta_{ml} = (b_{ml}/b_{\max})$); $\Theta = \|\theta_{ml}\|_{M \times L}$ – матрица коэффициентов длин каналов передачи данных между m-ми устройствами хранения и l-ми устройствами обработки;

14) q_{ml} – стоимость передачи единицы данных между на m-м устройством хранения и l-м устройством обработки ($m = \overline{1, M}; l = \overline{1, L}$); $Q = \|q_{ml}\|_{M \times L}$ – матрица стоимостей передачи единицы данных между m-ми устройствами хранения и l-ми ВУ.

Постановка задачи предполагает введение условия $\sum_{i=1}^{N} d_i \leq \sum_{m=1}^{M} v_m$, предусматривающего, что все данные будут распределены для хранения по устройствам. Необходимость построения расписания загрузки на устройства хранения в связи с этим отсутствует. Однако при распределении данных i-х типов по ВУ требуется построение расписаний их (данных) обработки на устройствах, так как на его основе осуществляется расчет стоимостных показателей, связанных с хранением.

В соответствии с иерархическим подходом [10, 11] к математическому моделированию и оптимизации процессов распределенного хранения и обработки данных выполнена декомпозиция обобщенной функции системы на совокупность иерархически упорядоченных подфункций, которые распределены по уровням следующим образом:

1) верхний уровень – моделирование и оптимизация распределенного хранения данных на устройствах;

2) нижний уровень – моделирование и оптимизация распределенной обработки данных на ВУ; также на нижнем уровне реализуется построение расписаний обработки на устройствах назначенных данных.

Построение в соответствии с выполненной декомпозицией математической модели процессов хранения и обработки данных предполагает введение в рассмотрение иерархически упорядоченных компонент, которые содержат следующие матрицы:

компонента верхнего уровня

– матрицу назначений $P = \|p_{im}\|_{N \times M}$, элементы которой $p_{im} = 1$, если данные i-го типа размещены для хранения на m-м устройстве, и $p_{im} = 0$, если данные i-го типа не размещены для хранения на m-м устройстве;

– матрицу $T^{mem} = \|t_{im}^{mem}\|_{N \times M}$ интервалов времени хранения данных i-х типов на m-х устройствах хранения ($i = \overline{1, N}; m = \overline{1, M}$); элемент матрицы $t_{im}^{mem} \neq 0$ в том случае, если для соответствующих индексов i и m элемент p_{im}матрицы P равен 1 ($p_{im} = 1$), элемент матрицы $t_{im}^{mem} = 0$ в том случае, если для соответствующих индексов i и m элемент p_{im} матрицы P равен 0 ($p_{im} = 0$);

компонента нижнего уровня

– матрицу назначений $R = \|r_{li}\|_{L \times N}$, элементы которой $r_{li} = 1$, если данные i-го типа назначены для обработки на l-м ВУ, и $r_{li} = 0$, если данные i-го типа не назначены для обработки на l-м устройстве;

– матрицу $T^0 = \left\|t_{il}^0\right\|_{N \times L}$моментов времени начала обработки данных i-х типов на l-х ВУ (элемент t_{il}^0 матрицы T^0 – это момент времени начала обработки данных i-го типа на l-м ВУ ($i = \overline{1, N}; l = \overline{1, L}$); $t_{il}^0 \neq 0$ в том случае, если для соответствующих индексов i и l элемент r_{li} матрицы R равен 1

($r_{li} = 1$), $t_{il}^0 = 0$ в том случае, если для соответствующих индексов i и l элемент r_{li} матрицы R равен 0 ($r_{li} = 0$)).

Элементы t_{il}^0 матрицы T^0 соответствуют определенным порядкам обработки данных i-х типов на l-х ВУ, формируемым с использованием разработанной эвристической процедуры. То есть матрица T^0 соответствует расписаниям обработки данных i-х типов на l-х ВУ.

В этом случае математическая модель процесса распределенного хранения и распределенной обработки данных имеет вид кортежей:

– *компонента верхнего уровня* – кортеж вида $[P, T^{mem}]$, соответствующий распределению данных для хранения по устройствам и характеризующий интервалы времени хранения их на этих устройствах;

– *компонента нижнего уровня* – кортеж вида $[R, T^0]$ соответствующий распределению данных на ВУ и характеризующий моменты времени начала их обработки на этих устройствах (иными словами – расписание распределенной обработки данных на ВУ).

Компоненты математической модели процесса распределенного хранения и обработки данных соответствуют решениям, оптимизируемым на каждом из уровней иерархии. На верхнем уровне оптимизируется решение по распределенному хранению данных и по интервалам времени, в течение которых данные определенных типов будут храниться на соответствующих устройствах. На нижнем уровне оптимизируется решение по распределенной обработке данных и по моментам времени начала их обработки на ВУ, которые соответствуют расписаниям выполнения операций с данными. Применение теоретико-игрового подхода предполагает назначение на верхнем уровне ведущего игрока, который выполняет оптимизацию решений по распределенному хранению данных, и назначение на нижнем уровне ведомого игрока, который выполняет оптимизацию решений по распределенной обработке данных (при условии, что расписания обработки данных, формируемые с использованием эвристической процедуры, однозначно соответствуют распределению обработки данных по ВУ).

Особенностью реализации теоретико-игрового подхода к иерархической оптимизации решений по распределенному хранению и обработке данных (особенностями взаимодействия игроков при оптимизации решений на уровнях игры) являются [12–14]:

1) передача решения по распределенному хранению данных с верхнего уровня на нижний с целью формирования на его основе решения по распределенной обработке данных;

2) формирование для полученного с верхнего уровня решения по распределенному хранению данных локально-оптимального решения по распределенной обработке данных на соответствующем ему нижнем уровне;

3) передача локально-оптимального решения по распределенной обработке данных с нижнего уровня на верхний с целью оценки оптимальности решения по распределенному хранению данных.

В соответствии с указанными особенностями взаимодействия игроков и введенными обозначениями для решений, оптимизируемым на каждом из уровней, модель иерархической игры представлена в следующем общем виде [12–14]:

1) верхний уровень:

$$\min_{[P,T^{mem}]\in N_1} f_1([P, T^{mem}], [R, T^0]*);$$

2) нижний уровень:

$$\min_{[R,T^0]\in N_2([P,T^{mem}])} f_2([P, T^{mem}], [R, T^0]),$$

где N_1, $N_2([P, T^{mem}])$ – множество решений по распределенному хранению данных и по их обработке, соответствующее сформированному решению $[P, T^{mem}]$ по распределенному хранению данных, полученному с верхнего уровня.

Построение математической модели иерархической игры реализуется в предположении, что на верхнем уровне оптимизация решений выполняется с учетом внешней цели функционирования системы, которая требует минимизации финансовых затрат на хранение, передачу и обработку данных (минимизацию стоимости использования ресурсов). Оптимизация решений по распределенной обработке данных на нижнем уровне выполняется с точки зрения внутренней цели функционирования системы, предусматривающей минимизацию длительностей передачи и обработки данных. На основе указанных особенностей формирование критериев на уровнях иерархической игры оптимизации решений по распределенному хранению и обработке данных (при учете передачи данных между устройствами) выполняется в соответствии с рассматриваемыми ниже рассуждениями.

Интервал времени передачи данных одного i-го типа между m-м устройством, на котором реализуется их хранение, и l-м ВУ, на котором реализуется их обработка, определяется следующим образом:

$$\sum_{m=1}^{M}\sum_{l=1}^{L} p_{im} r_{li} \frac{d_i}{c_{ml}}.$$

В представленном выражении не учитывается длина канала связи, который используется для передачи данных. В рассмотрение введен коэффициент θ_{ml}, выступающий в роли веса, позволяющего учесть длины каналов связи в формируемых

оценках критерия при сравнении решений по распределенной обработке данных.

В этом случае оценка времени передачи данных одного i-го типа между m-м устройством, на котором реализуется их хранение, и l-м ВУ, на котором реализуется их обработка (с учетом веса, характеризующего длину канала связи), определяется выражением:

$$\sum_{m=1}^{M}\sum_{l=1}^{L} p_{im} r_{li} \frac{d_i}{c_{ml}} \theta_{ml}.$$

Тогда обобщенная оценка времени передачи данных всех n типов между устройствами, на которых реализуется их хранение, и устройствами, на которых реализуется их обработка (с учетом весов, соответствующих длинам каналов передачи данных), определяется выражением:

$$\sum_{i=1}^{N}\sum_{m=1}^{M}\sum_{l=1}^{L} p_{im} r_{li} \frac{d_i}{c_{ml}} \theta_{ml}.$$

Интервал времени выполнения вычислений с данными i-х типов, обработка которых назначена на одном l-м устройстве, определяется следующим образом:

$$\sum_{i=1}^{N} r_{li} t_{il}.$$

При условии, что обработку данных на всех устройствах требуется распределить равномерно (равномерная загрузка ВУ обработкой данных), интервал времени, соответствующий окончанию обработки данных всех N типов на всех параллельно функционирующих устройствах, определяется выражением:

$$\max_{l=\overline{1,L}} \sum_{i=1}^{N} r_{li} t_{il}.$$

На нижнем уровне иерархической игры необходимо минимизировать общие временные затраты на передачу данных между устройствами хранения и обработки, а также временные затраты на обработку данных (с целью минимизации общего времени использования ресурсов обработки и передачи данных).

Указанные характеристики процесса распределенной обработки данных определяются выражением следующего вида:

$$\sum_{i=1}^{N}\sum_{m=1}^{M}\sum_{l=1}^{L} p_{im} r_{li} \frac{d_i}{c_{ml}} \theta_{ml} + \max_{l=\overline{1,L}} \sum_{i=1}^{N} r_{li} t_{il}. \qquad (1)$$

Выражение (1) используется в качестве критерия оптимальности решений по распределенной обработке данных всех N типов на L параллельно функционирующих устройствах.

Определение интервалов времени t_{im}^{mem} хранения данных i-х типов на m-х устройствах ($i = \overline{1,N}; m = \overline{1,M}$) реализуется в предположении, что все данные N типов распределяются по устройствам хранения одновременно в момент времени, равный 0. Тогда интервал времени хранения данных i-го типа на m-м устройстве определяется с учетом момента времени t_{il}^{0} начала обработки данных на l-м ВУ и интервала времени передачи всего объема данных этого типа (в количестве d_i) с m-го устройства на l-е устройство. То есть момент времени окончания хранения данных i-го типа на m-м устройстве (и, соответственно, интервал времени t_{im}^{mem} их хранения на этом устройстве) определяется как разность между моментом времени начала обработки данных на l-м устройстве и интервалом времени передачи данных по каналу, соединяющему m-е устройство (где хранятся данные) и l-е устройство (где они обрабатываются).

Интервал времени передачи данных между m-м устройством хранения и l-м устройством обработки определяется как сумма интервала времени выставления данных в канал в заданном их объеме d_i (при известной пропускной способности c_{ml} канала, по которому передаются данные) и интервала времени передачи данных по каналу. С целью определения интервала времени передачи данных по каналу в рассмотрение введена константа S_{tr}, соответствующая скорости распространения сигнала в канале ($S_{tr} = 2 * 10^5$км/с).

Тогда для данных i-го типа, размещенных для хранения на m-м устройстве и обрабатываемых на l-м устройстве, время их передачи по каналу будет определено следующим образом:

$$\sum_{l=1}^{L} p_{im} \cdot r_{li} \cdot \frac{b_{ml}}{S_{tr}},$$

где b_{ml} – длина канала между m-м устройством (где данные размещены для хранения) и l-м устройством, на котором данные обрабатываются.

Интервал времени выставления данных i-го типа, размещенных для хранения на m-м устройстве, в канал, соединяющий это устройство и l-е устройство, на котором они обрабатываются, определяется выражением вида:

$$\sum_{l=1}^{L} p_{im} \cdot r_{li} \cdot \frac{d_i}{c_{ml}}.$$

Тогда общий интервал времени выставления данных i-го типа (количество которых равно d_i) в канал с пропускной способностью c_{ml} и передачи этих данных по каналу, характеризуемому длиной b_{ml}, определяется следующим образом (при зафиксированном значении типа данных i и индексе m хранилища, на котором они находятся):

$$\sum_{l=1}^{L} p_{im} r_{li} \left(\frac{d_i}{c_{ml}} + \frac{b_{ml}}{S_{tr}} \right).$$

Значение t_{il}^0 для данных i-го типа, обрабатываемых на l-м устройстве, определяется в соответствии с расписанием, формируемом с использованием эвристической процедуры.

Тогда интервал времени t_{im}^{mem} хранения данных i-го типа на m-м устройстве определяется следующим образом (значения идентификатора i типа данных и идентификатора устройства хранения m зафиксированы):

$$t_{im}^{mem} = \sum_{l=1}^{L} \left(t_{il}^0 r_{li} - p_{im} r_{li} \left(\frac{d_i}{c_{ml}} + \frac{b_{ml}}{S_{tr}} \right) \right),$$

где $i = \overline{1, N}, m = \overline{1, M}$.

С целью синтеза критерия оптимальности решений на верхнем уровне иерархической игры введена в рассмотрение матрица $P' = \|p'_{mi}\|_{M \times N}$, элементы которой определяются следующим образом: $P' = P^T$ (где T – символ операции транспонирования матрицы P). То есть $p'_{mi} = p_{im}$ $(m = \overline{1, M}; i = \overline{1, N})$.

Тогда для данных i-го типа, хранящихся на m-м устройстве, стоимость хранения единицы данных в течение интервала времени t_{im}^{mem} определяется выражением вида:

$$\sum_{m=1}^{M} t_{im}^{mem} \cdot p'_{mi} s_m.$$

В силу истинности условия:

$$\sum_{i=1}^{N} d_i \le \sum_{m=1}^{M} v_m$$

все данные будут распределены для хранения по устройствам.

Тогда стоимость хранения всех данных n типов на устройствах определяется выражением вида:

$$\sum_{i=1}^{N} d_i \left(\sum_{m=1}^{M} t_{im}^{mem} p'_{mi} s_m \right).$$

где d_i – это количество данных i-х типов, хранящихся на m-х устройствах.

Стоимость выполнения единицы вычислительных операций на l-ом устройстве обозначена через w_l, а длительность выполнения всех операций на l-м устройстве с назначенными на него данными определяется выражением:

$$\sum_{i=1}^{N} r_{li} t_{il}.$$

Тогда стоимость выполнения всех вычислительных операций на отдельном l-м устройстве определяется выражением:

$$w_l \sum_{i=1}^{N} r_{li} t_{il},$$

а стоимость реализации вычислительных операций с данными на всех L устройствах обработки:

$$\sum_{l=1}^{L} w_l \sum_{i=1}^{N} r_{li} t_{il}.$$

Аналогичным образом затраты на передачу единицы данных разных i-х типов между некоторым m-м устройством хранения, на котором они размещены, и некоторым l-м устройством обработки, на котором они обрабатываются, определяются выражением:

$$\sum_{i=1}^{N} p_{im} r_{li} q_{ml},$$

а затраты на передачу всего объема d_i данных i-х типов между m-м устройством хранения и l-м устройством обработки определяется выражением:

$$\sum_{i=1}^{N} p_{im} r_{li} d_i q_{ml}.$$

Выражение для определения суммарных затрат на передачу данных i-х типов $(i = \overline{1, N})$ от разных m-х устройств хранения на различные l-е устройства ~~для~~ обработки имеет вид:

$$\sum_{i=1}^{N} \sum_{m=1}^{M} \sum_{l=1}^{L} p_{im} r_{li} d_i q_{ml}.$$

Количество использованного ресурса на некотором m-м устройстве для хранения данных i-го типа определяется выражением:

$$\sum_{i=1}^{N} p'_{mi} d_i,$$

где p'_{mi} – элемент матрицы P' $(P' = P^T)$.

Количество неиспользованного ресурса на m-м устройстве хранения определяется выражением:

$$v_m - \sum_{i=1}^{N} p'_{mi} d_i.$$

Тогда размер штрафа за неиспользованный ресурс хранения данных на m-м устройстве определяется выражением:

$$s_m \left(v_m - \sum_{i=1}^{N} p'_{mi} d_i \right).$$

Итоговое выражение для определения суммарных штрафов за неиспользуемые ресурсы на всех устройствах хранения имеет вид:

$$\sum_{m=1}^{M} s_m \left(v_m - \sum_{i=1}^{N} p'_{mi} d_i \right).$$

Выражение для суммарных затрат на распределенное хранение и обработку данных, передачу данных между устройствами хранения и обработки при учете штрафов за неиспользование ограниченных ресурсов их хранения, полученное на основе синтезированных выше выражений, имеет вид:

$$\sum_{i=1}^{N} d_i \left(\sum_{m=1}^{M} t_{im}^{mem} p'_{mi} s_m \right) + \sum_{l=1}^{L} w_l \sum_{i=1}^{N} r_{li} t_{il} + \sum_{i=1}^{N} \sum_{m=1}^{M} \sum_{l=1}^{L} p_{im} r_{li} d_i q_{ml} + \sum_{m=1}^{M} s_m \left(v_m - \sum_{i=1}^{N} p'_{mi} d_i \right). \quad (2)$$

С учетом выражений (1, 2) для критериев оптимальности решений по распределенному хранению и обработке данных получена математическая модель иерархической игры оптимизации указанных решений в следующем виде:

1) верхний уровень – минимизация затрат на хранение, передачу, обработку данных и штрафов за не полное использование ограниченных ресурсов:

$$\min \ f_1 \ , \quad (3)$$

где

$$f_1 = \sum_{i=1}^{N} d_i \left(\sum_{m=1}^{M} t_{im}^{mem} p'_{mi} s_m \right) + \sum_{l=1}^{L} w_l \sum_{i=1}^{N} r_{li} t_{il} + \sum_{i=1}^{N} \sum_{m=1}^{M} \sum_{l=1}^{L} p_{im} r_{li} d_i q_{ml} + \sum_{m=1}^{M} s_m \left(v_m - \sum_{i=1}^{N} p'_{mi} d_i \right);$$

2) нижний уровень – минимизация времени передачи данных и времени обработки:

$$\min \ f_2 \ , \quad (4)$$

где

$$f_2 = \sum_{i=1}^{N} \sum_{m=1}^{M} \sum_{l=1}^{L} p_{im} r_{li} \frac{d_i}{c_{ml}} \theta_{ml} + \max_{l=\overline{1,L}} \sum_{i=1}^{N} r_{li} t_{il};$$

3) ограничения на множества допустимых решений на верхнем и нижнем уровне:

$$\sum_{m=1}^{M} p_{im} = 1 \ (i = \overline{1,N}), \quad (5)$$

$$\sum_{l=1}^{L} r_{li} = 1 \ (i = \overline{1,N}), \quad (6)$$

$$\sum_{i=1}^{N} d_i \cdot p_{im} \le v_m \ (m = \overline{1,M}). \quad (7)$$

Ограничение (5) предусматривает, что данные i-го типа размещаются только на одном из M устройств их хранения. Ограничение (6) предусматривает, что данные i-го типа назначаются только на одно из LВУ для их обработки. Ограничение (7) предусматривает, что количество данных разных типов, хранящихся на m-м устройстве, не превышает допустимого (заданного) объема хранилища.

Таким образом, математическая модель иерархической игры вида (3–7) представляет собой способ оптимизации решений по распределенному хранению и распределенной обработке данных, реализуемый в двухуровневой системе.

2. Генетический алгоритм оптимизации решений по распределенному хранению и обработке данных

На нижнем уровне иерархии системы оптимизации решений по распределенному хранению и обработке данных обеспечивается достижение внутренней цели функционирования, предусматривающей эффективное использование ресурсов (минимизацию временных затрат). Поэтому распределение обработки данных по ВУ реализуется с учетом решения по распределенному хранению данных таким образом, чтобы обеспечить минимизацию времени использования каналов передачи данных. Для минимизации времени хранения данных на устройствах расписание (порядок) их обработки формируется в соответствии с эвристическим алгоритмом, предусматривающим, что среди всех данных, закрепленных для выполнения на соответствующем ВУ, в первую очередь обрабатываются данные большого объема. Тем самым обеспечивается минимизация времени использования ресурса хранения. Минимизация использования ресурсов обработки и передачи данных гарантируется видом сформированного критерия.

С целью оптимизации решений по распределенному хранению и обработке данных применен аппарат генетических алгоритмов [15, 16]. Реализация генетических алгоритмов для оптимизации решений на соответствующих уровнях иерархии предусматривает первоначальную разработку способа их (решений) кодирования и последующую разработку генетических операторов, которые позволяют получить локально-оптимальные решения на соответствующих уровнях иерархической игры.

Кодирование решений по распределенной обработке данных с целью их оптимизации на нижнем уровне иерархической игры выполняется следующим образом. Одному решению по распределенной обработке данных соответствует хромосома, состоящая из n генов (где n – количество типов данных, обрабатываемых в системе). Каждый i-й ген $(i = \overline{1,N})$ соответствует закреплению данных i-го типа для обработки за одним из L ВУ. Обозначим значение i-го гена через l_i $(i = \overline{1,N})$, тогда при формировании начального решения (хромосомы) выполняется инициализация $l_i = l$, где l – номер ВУ, на котором выполняется обработка данных i-го типа $(l = \overline{1,L})$.

Для кодирования решений по распределенному хранению данных с целью их оптимизации на верхнем уровне иерархической игры каждому i-му гену $(i = \overline{1,N})$ ставится в соответствие номер устройства m $(m = \overline{1,M})$, на котором размещаются данные i-го типа. Значение i-го гена $(i = \overline{1,N})$ обозначим через m_i, тогда при формировании некоторого решения (хромосомы) выполняется инициализация $m_i = m$, где m – номер устройства $(m = \overline{1,M})$, на котором выполняется хранение данных i-го типа $(i = \overline{1,N})$.

Популяция (множество) хромосом, соответствующих решениям по размещению данных на устройствах хранения, имеет вид:

$$\mathrm{H}_m = \{h_{m_u} | u = \overline{1,U}\},$$

где U – количество хромосом в популяции; h_{m_u} – u-я хромосома, представленная в виде вектора $h_{m_u} = (m_1^u, m_2^u, .., m_n^u)$; m_i^u – значение i-го гена $(i = \overline{1,N})$ в u-й хромосоме $(u = \overline{1,U})$, при условии, что $m_i^u \in \{1,2,\dots,M\}$.

Популяция (множество) хромосом, соответствующих решениям по назначению обработки данных на соответствующих ВУ, имеет вид:

$$\mathrm{H}_l = \{h_{l_u} | u = \overline{1,U}\},$$

где h_{l_u} – u-я хромосома, представленная в виде вектора $h_{l_u} = (l_1^u, l_2^u, .., l_n^u)$; l_i^u – значение i-го гена $(i = \overline{1,N})$ в u-й хромосоме $(u = \overline{1,U})$, такое, что $l_i^u \in \{1,2,\dots,L\}$.

В соответствии с представленным способом кодирования решений формируется начальная популяция путем генерации значений соответствующих генов разных хромосом (при условии выполнения ограничения (7)). Дальнейшая оптимизация решений предусматривает их изменение на каждом из уровней иерархии в иерархической игре в соответствии с разработанными генетическими операторами. Для поиска локально оптимальных решений по распределенному хранению и обработке данных использованы следующие генетические операторы:

– оператор селекции с целью выбора родительских хромосом для их последующего скрещивания;

– оператор кроссоверинга (скрещивания), позволяющий получать новые хромосомы (решения) путем наследования свойств родительских хромосом (решений);

– оператор мутации, позволяющий трансформировать решения, полученные в результате скрещивания, случайным образом.

В качестве оператора селекции родительских хромосом используется метод, основанный на принципе колеса рулетки. Так как на каждом из уровней в иерархический игре реализуется минимизация соответствующих этим уровням критериев, поэтому разработан способ определения сектора колеса рулетки, соответствующего рассматриваемому решению (способ определения величины подинтервала в интервале [0;1]), который ставится в соответствие рассматриваемой хромосоме (решению) при определении возможности ее скрещивания с другими хромосомами.

Для каждой хромосомы $h_{m_u} \in H_m$ (аналогично для хромосомы $h_{l_u} \in H_l$) первоначально определяется ее оценка $O_{m_u}^1$, используемая в дальнейшем при вычислении величины сектора, в соответствии с выражением вида:

$$O_{m_u}^1 = \frac{f_1^u}{\sum_{u=1}^{U} f_1^u},$$

для хромосом, соответствующих решениям по распределению данных на ВУ для обработки:

$$O_{l_u}^1 = \frac{f_2^u}{\sum_{u=1}^{U} f_2^u},$$

где f_1^u и f_2^u – значения критериев f_1 и f_2 на верхнем и нижнем уровнях иерархической игры для некоторых u-х решений (u-х хромосом, входящих в популяции)).

На основе значений $O_{m_u}^1$ (для решений на верхнем уровне) и $O_{l_u}^1$ (для решений на нижнем уровне) определяются значения для оценок $O_{m_u}^2$ и $O_{l_u}^2$ u-х хромосом $(u = \overline{1,U})$ в соответствии с выражениями вида:

$$O_{m_u}^2 = \frac{1}{1 + O_{m_u}^1};\ O_{l_u}^2 = \frac{1}{1 + O_{l_u}^1}.$$

Оценка P_{m_u} и и P_{l_u} величины сектора колеса рулетки для u-х хромосом $(u = \overline{1,U})$ на верхнем и нижнем уровнях иерархической игры определяются в соответствии с выражениями вида:

$$P_{m_u} = \frac{O_{m_u}^2}{\sum_{u=1}^{U} O_{m_u}^2};\ P_{l_u} = \frac{O_{l_u}^2}{\sum_{u=1}^{U} O_{l_u}^2}.$$

Вычисленные с использованием полученных выражений значения величин сектора колеса рулетки используются для определения размеров подинтервалов в интервале [0;1], закрепляемых за соответствующими хромосомами. Генерация случайных чисел и определение их принадлежности соответствующим подинтервалам позволяют идентифицировать родительские хромосомы, используемые для скрещивания. После формирования множества родительских хромосом, используемых для скрещивания, в нем необходимо идентифицировать пары, к которым будет применен оператор кроссоверинга. Способами выбора пар родительских хромосом в соответствующем множестве для их непосредственного скрещивания (применения оператора кроссоверинга) являются [15, 16]:

а) случайный выбор – пары родительских хромосом выбираются случайным образом;

б) селективный выбор – выбираются такие хромосомы, которым соответствуют значения критерия оптимальности решений выше среднего значения для всей популяции;

в) инбридинг, при котором первая родительская хромосома выбирается случайным образом, а вторая – исходя из максимальной похожести на первую;

г) аутбридинг, при котором первая родительская хромосома выбирается случайно, а вторая – исходя из максимального отличия от первой.

С целью исключения возможности преждевременного попадания в «ловушку» локального экстремума для выбора пар родительских хромосом использован генотипный аутбридинг.

Идентификация различия между текущей рассматриваемой хромосомой $h_{m_u} \in H_m (h_{l_u} \in H_l$ в популяции $H_l)$, выбираемой случайным образом, и хромосомой $h_{m_j} \in H_m$, с которой будет выполняться скрещивание, определяется в соответствии с условием следующего вида:

$$\max_j \sum_{i=1}^{N} \left| m_i^u - m_i^j \right| ; \ \max_j \sum_{i=1}^{N} \left| l_i^u - l_i^j \right|,$$

где $j = \overline{1, u-1} \& j = \overline{u+1, U})$.

Представленное условие определяет максимальное различие в генах в хромосоме, выбранной случайно, и в хромосоме, с ней скрещиваемой.

Для выбранных таким образом пар хромосом случайным образом определяются 2 точки скрещивания (номера генов в хромосомах, являющихся граничными при реализации скрещивания). Индексы (номера) генов, являющихся границами участков хромосом при скрещивании обозначены через q_1 (первая точка) и q_2 (вторая точка). Таким образом, в генах с определенными таким образом номерами реализуется двухточечное скрещивание пары родительских хромосом.

Если родительские хромосомы имеют вид:

$$h_{m_u} = (m_1^u, m_2^u, \dots, m_{q_1-1}^u, m_{q_1}^u, m_{q_1+1}^u, \dots, \dots m_{q_2-1}^u, m_{q_2}^u, m_{q_2+1}^u, \dots, m_n^u)$$

и

$$h_{m_j} = (m_1^j, m_2^j, \dots, m_{q_1-1}^j, m_{q_1}^j, m_{q_1+1}^j, \dots, \dots, m_{q_2-1}^j, m_{q_2}^j, m_{q_2+1}^j, \dots, m_n^j),$$

то в результате реализации оператора двухточечного кроссоверинга пара дочерних хромосом будет иметь вид:

$$h'_{m_u} = (m_1^u, m_2^u, \dots, m_{q_1-1}^u, m_{q_1}^j, m_{q_1+1}^j, \dots, \dots, m_{q_1-1}^j, m_{q_1}^u, m_{q_1+1}^u, \dots, m_n^u)$$

и

$$h'_{m_j} = (m_1^j, m_2^j, \dots, m_{q_1-1}^j, m_{q_1}^u, m_{q_1+1}^u, \dots, \dots, m_{q_1-1}^u, m_{q_1}^j, m_{q_1+1}^j, \dots, m_n^j).$$

Аналогичные рассуждения применены при скрещивании хромосом h_{l_u} и h_{l_j} на нижнем уровне иерархии игры при оптимизации решений по назначению ВУ для обработки данных n типов.

Полученные описанным способом дочерние хромосомы подвергаются на следующем этапе воздействию оператора мутации. Стохастический характер оператора мутации предусматривает, что каждый ген будет подвержен ее (мутации) воздействию с определенной вероятностью (обозначенной как $P_м$). Воздействие оператора мутации на ген предусматривает либо увеличение на 1 его значения, либо уменьшения на 1 его значения. Характер операции (увеличение либо уменьшение значений генов) также определяется стохастически.

Синтезированный генетический алгоритм используется для поиска локально оптимального решения по закреплению обработки данных i-х типов ($i = \overline{1, N}$) за l-ми ВУ ($l = \overline{1, L}$) (лучшей хромосомы в сформированной финальной популяции, сгенерированной для текущего решения по распределенному хранению данных). То есть это локально оптимальное решение соответствует текущему решению по распределенному хранению данных, полученному с верхнего уровня. Оно передается на верхний уровень с целью оценки оптимальности решения по распределенному хранению данных. Для решений по распределенному хранению данных на верхнем уровне (хромосом, входящих в популяцию, рассматриваемую на верхнем уровне) реализуется повторное формирование новых решений с использованием рассмотренных выше генетических операторов (формирование новой популяции хромосом, соответствующих решениям по распределенному хранению данных).

Поиск локально оптимальных решений по распределенному хранению данных и их распреде-

ленной обработке реализуется для заданного количества поколений популяций. После чего в финальной популяции на верхнем уровне выбирается лучшее (локально оптимальное) решение (с минимальным значением критерия на этом уровне), а на нижнем уровне идентифицируется лучшее решение в популяции, соответствующей этому локально оптимальному решению на верхнем уровне.

Оптимизация решений по распределенному хранению и обработке данных (на верхнем и нижнем уровне иерархической игры, соответственно) выполняется при следующих параметрах генетических алгоритмов: общее количество хромосом в популяции U = 60; разрыв поколений T = 0,5 (количество наиболее приспособленных хромосом, участвующих в селекции), количество поколений равно 30. Вероятность мутации генов определяется по формуле $P_м = 1/(10 \cdot N)$, где N – количество типов данных, распределенное хранение и обработка которых реализуется в системе.

3. Эвристический алгоритм построения расписаний обработки данных на ВУ

Поиск локально оптимальных решений по распределенному хранению и обработке данных реализуется при условии минимизации стоимости использования ресурсов (в том числе минимизации стоимости передачи данных по каналам, связывающим устройства хранения и обработки). С целью минимизации интервалов времени хранения данных на m-х устройствах (и как следствие – минимизации стоимости хранения) предложен эвристический алгоритм, позволяющий формировать расписания обработки данных на назначенных для этого ВУ. Таким образом, построение расписаний обработки данных на ВУ, на которых она назначена, реализуется с учетом требования минимизации времени нахождения данных в хранилищах, что обеспечивает минимизацию стоимости хранения. Эвристический алгоритм построения расписаний предусматривает, что среди всех назначенных на устройство для обработки данных в первую очередь будут обрабатываться те, размер которых является максимальным (то есть обработка данных на устройстве осуществляется в порядке не убывания их объемов $d_i (i = \overline{1,N})$).

Для обоснования эвристической процедуры формирования порядков реализации распределенной обработки данных на каждом из ВУ в рассмотрение введены:

1) множества $N^l (l = \overline{1,L})$ типов данных, назначенных для выполнения указанной операции на эти устройства (множества формируются в соответствии с видом матрицы R с учетом значений ее элементов $r_{li} = 1$ в каждой l-й строке $(l = \overline{1,L})$);

2) переменные $Pred^l$ $(l = \overline{1,L})$, предназначенные для хранения идентификатора типа данных, которые рассматривались на предшествующей итерации алгоритма и для которых было вычислено значение t_{il}^0 элемента матрицы T^0, соответствующей расписаниям обработки данных на l-х приборах (первоначально переменные инициализируются значением 0).

Определение значений t_{il}^0 $(i = \overline{1,N}, l = \overline{1,L})$ моментов времени начала выполнения операций с данными i-х типов на l-х ВУ (то есть построение расписаний обработки данных на l-х устройствах) реализуется в предположении, что к моменту времени начала интерпретации расписания данные всех N типов уже распределены по устройствам хранения. В том случае, если данные некоторых i-х типов занимают $(j = 1)$-е позиции в последовательностях их обработки на l-х приборах (значения переменных $Pred^l = 0 (l = \overline{1,L})$), тогда для них предполагается, что $t_{im}^{mem} = 0$. В этом случае момент времени начала обработки данных определяется на основе интервала времени их передачи между m-м устройством, где они размещены, и l-м ВУ, на котором они обрабатываются (определяется интервалом времени выставления данных в канал и интервалом времени передачи данных по каналу).

Вычисление значения t_{il}^0 для данных i-го типа в $(j = 1)$-й позиции их обработки на l-м приборе реализуется в соответствии с выражением:

$$t_{il}^0 = \sum_{m=1}^{M} p_{im} r_{li} \left(\frac{d_i}{c_{ml}} + \frac{b_{ml}}{S_{tr}} \right). \quad (8)$$

В том случае, если для данных i-го типа их позиция в последовательности обработки на l-ом приборе $j \neq 1$, то значение t_{il}^0 определяется выражением:

$$t_{il}^0 = t_{i'l}^0 + \sum_{l=1}^{L} t_{i'l} r_{li'}, \quad (9)$$

где i' – тип данных, которые занимают в последовательности их обработки на l-м приборе предшествующую $(j-1)$-ю позицию по отношению к j-й позиции данных рассматриваемого i-го типа (номер типа данных использован на предыдущей итерации алгоритма построения расписания для инициализации переменной $Pred^l$). Перед началом интерпретации процедуры построения расписаний и определения значений t_{il}^0 для данных i-х типов, закрепленных за соответствующими приборами, элементы матрицы T^0 инициализированы значением 0.

Эвристический алгоритм построения расписания обработки данных, назначенных на l-е устройство, а также определения моментов времени t_{il}^0

начала обработки данных i-х типов на этом устройстве, имеет следующий порядок шагов.

Шаг 1. Переменную $\mathrm{Pr}ed^l$ инициализировать значением 0.

Шаг 2. Определить в множестве N^l тип данных i', которому соответствует условие: $\max_i(d_i | \ i \in N^l)$.

Шаг 3. Если $\mathrm{Pr}ed^l = 0$, то для рассматриваемого i'-го типа данных определить значение $t^0_{i'l}$ с использованием выражения (8); если $\mathrm{Pr}ed^l \neq 0$, то для рассматриваемого i'-го типа данных определить значение $t^0_{i'l}$ с использованием выражения (9), с учетом данных i-го типа, идентификатором которых инициализирована переменная $\mathrm{Pr}ed^l$ на предыдущей итерации алгоритма, с использованием значения t^0_{il}, определяемого для данных этого типа в матрице T^0).

Шаг 4. Модифицировать множество N^l и значение переменной $\mathrm{Pr}ed^l$: $N^l = N^l \backslash \{i'\}$; $\mathrm{Pr}ed^l = i'$.

Шаг 5. Если $N^l \neq \emptyset$, то перейти на пункт 2; если $N^l = \emptyset$, то перейти на пункт 6.

Шаг 6. Останов алгоритма.

Результатом интерпретации рассмотренного алгоритма являются последовательности выполнения заданий на l-х приборах, которым соответствует матрица T^0 моментов времени t^0_{il} начала обработки данных i-х типов на этих приборах. Матрица T^0 совместно с матрицей назначений R передается на верхний уровень с целью вычисления на их основе оценки критерия f_1 текущего рассматриваемого решения $[P, T^{mem}]$ по распределению данных по хранилищам.

4. Анализ эффективности применения метода многоуровневой оптимизации решений по распределенному хранению и обработке данных

С целью исследования эффективности применения предложенного метода многоуровневой оптимизации решений по распределенном хранению и обработке данных значения параметров задачи задаются в соответствии с следующими обозначениями:

$\min(d_i)$ – минимальный объем данных i-х типов ($i = \overline{1, N}$), обрабатываемых в системе;

$\max(d_i)$ – максимальный объем данных i-х типов ($i = \overline{1, N}$), обрабатываемых в системе;

$\max(d_i)/\min(d_i)$ – отношение максимального объема данных i-х типов ($i = \overline{1, N}$) к минимальному объему данных i-х типов ($i = \overline{1, N}$), обрабатываемых в системе, которое определяет неоднородность объемов данных различных типов, распределяемых по устройствам хранения;

$\min(t_{il})$ – минимальная длительность обработки данных i-х типов ($i = \overline{1, N}$) на l-х ВУ ($l = \overline{1, L}$);

$\max(t_{il})$ – максимальная длительность обработки данных i-х типов ($i = \overline{1, N}$) на l-х ВУ ($l = \overline{1, L}$);

$\max(t_{il})/\min(t_{il})$ – отношение максимальной длительности обработки данных i-х типов ($i = \overline{1, N}$) на l-х ВУ ($l = \overline{1, L}$) к минимальной, определяющее неоднородность длительностей обработки данных на различных ВУ.

При проведении исследований значения $\min(d_i)$ и $\min(t_{il})$ задавались равными 10. Также при проведении исследований значения отношений $\max(d_i)/\min(d_i)$ и $\max(t_{il})/\min(t_{il})$ задавались равными: 1, 2, 4, 8, 16. Исследования проводились при фиксированных значениях длин каналов передачи данных между устройствами хранения и обработки, а также их пропускных способностей.

С целью анализа эффективности применения метода многоуровневой оптимизации решений по распределенному хранению и обработке данных использован эвристический алгоритм SIM упаковки в контейнеры [17]; метод формируется путем оптимизации с использованием рассмотренного генетического алгоритма. Эффективность применения метода многоуровневой оптимизации характеризуется степенью снижения затрат на хранение, обработку, передачу данных и штрафов за неиспользованные ресурсы для решений, полученного с использованием эвристического алгоритма SIM, и для решений, полученных путем оптимизации.

Снижение суммарных затрат при использовании метода многоуровневой оптимизации оценивалась путем сравнения:

1) значения критерия f_1 на верхнем уровне, обозначенного как f_1^{SIM}, полученного для решения, сформированного с использованием эвристического алгоритма SIM упаковки в контейнеры [17], и соответствующего ему локально оптимального решения по распределенной обработке данных;

2) значения критерия f_1 на верхнем уровне, полученного после многоуровневой оптимизации (обозначенного как f_1^o).

Снижение затрат на хранение, обработку, передачу данных и штрафов при оптимизации решений определено в соответствии с выражением:

$$\frac{f_1^{SIM} - f_1^o}{f_1^{SIM}}.$$

Снижение суммарных затрат на хранение, обработку, передачу данных и штрафов, получаемое при использовании метода многоуровневой оптимизации, представлено на рисунке 1 (слева M = 5, L = 5; справа M = 10, L = 10).

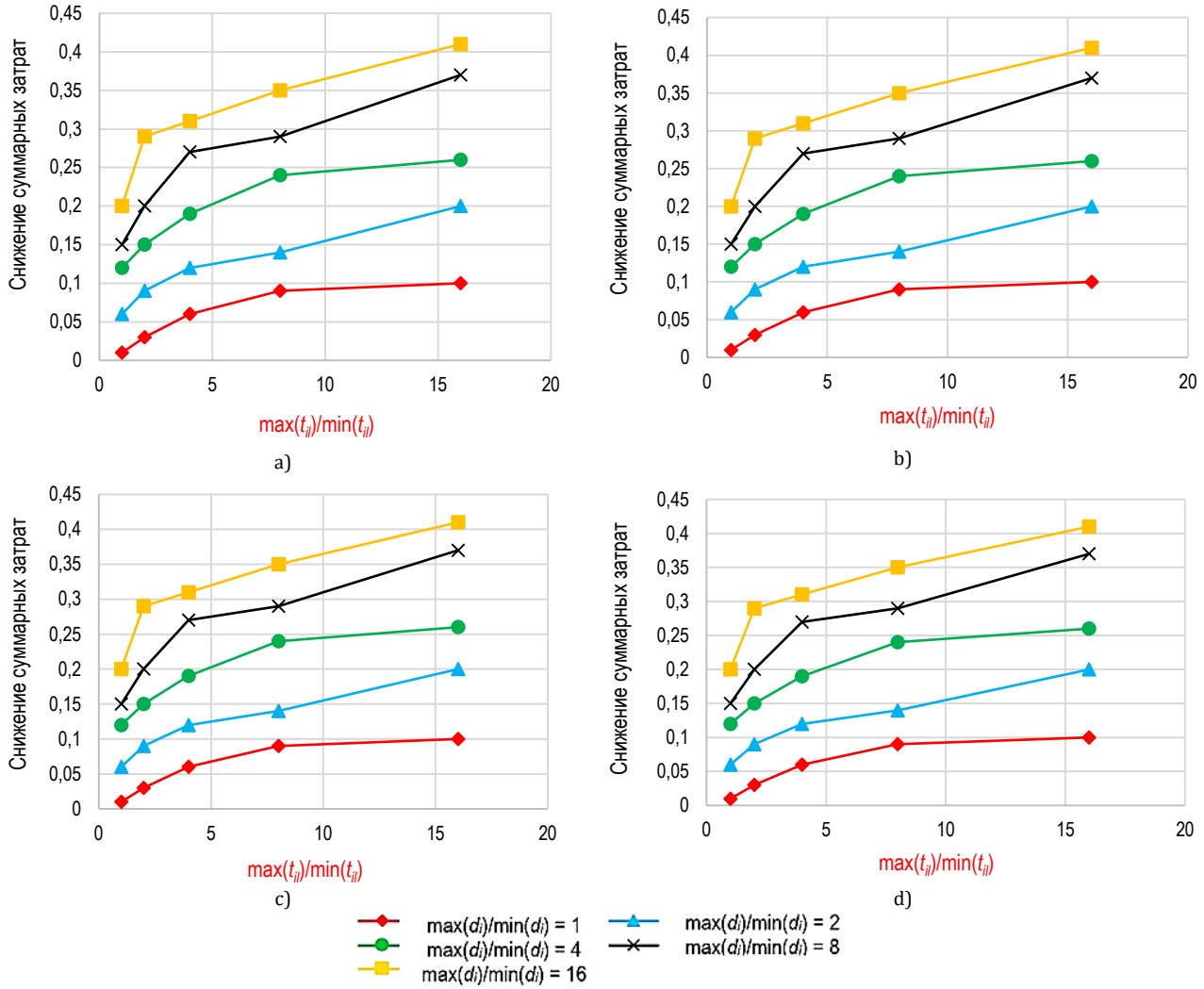


Рис. 1. Снижение суммарных затрат на хранение, обработку, передачу данных и штрафов для решений, полученных с использованием алгоритма SIM и метода многоуровневой оптимизации для n = 50 (a, c) и для n = 100 (b, d)

Fig. 1. Reduction of Total Costs for Storage, Processing, Data Transmission and Penalties for Solutions Obtained Using the SIM Algorithm and the Multilevel Optimization Method for n = 50 (a, c) and for n = 100 (b, d)

Анализ результатов исследований позволил сделать следующие выводы, касающиеся применения предложенного метода многоуровневой оптимизации решений по распределенному хранению и обработке данных:

1) увеличение количества устройств хранения M и обработки данных L (при неизменном количестве типов данных n) обуславливает снижение эффективности применении разработанного метода;

2) увеличение количества типов данных n, для которых реализуется распределенное хранение и обработка (при неизменном количестве ВУ) обуславливает увеличение эффективности применения разработанного метода;

3) увеличение неоднородности объемов хранимых данных (значения $\max(d_i)/\min(d_i)$) обуславливает более значительный рост эффективности применения метода по сравнению с увеличением неоднородности длительностей обработки данных на устройствах (значения $\max(t_{il})/\min(t_{il})$).

Заключение

В работе решена задача математического моделирования и оптимизации процессов распределенного хранения и обработки данных с учетом ограничений на ресурсы и взаимодействия между устройствами. Разработана иерархическая модель процесса распределенного хранения и обработки данных, представляющая собой совокупность компонент на соответствующих уровнях. Компонента модели верхнего уровня соответствует составляющей процесса, связанной с распределенным хранением данных, компонента нижнего уровня – с составляющей процесса, связанной с распределенной их обработкой вычислительными устройствами.

Для определения эффективного размещения данных с целью их распределенного хранения и обработки применена декомпозиция обобщенной задачи оптимизации на совокупность иерархически упорядоченных подзадач, для каждой из кото-

рых на соответствующем ей уровне идентифицируется локально-оптимальное решение. На основе определенной таким образом совокупности иерархически упорядоченных подзадач разработана математическая модель иерархической игры идентификации локально оптимальных решений на соответствующих уровнях. Предложенная модель игры позволяет реализовать совместную оптимизацию размещения данных при их распределенном хранении (на верхнем уровне) и их закрепления за вычислительными устройствами для распределенной обработки (на нижнем уровне). Критерий на верхнем уровне соответствует суммарным затратам на хранение, обработку, передачу данных, а также штрафы за неполное использование ограниченных ресурсов хранения. Критерий на нижнем уровне соответствует суммарным длительностям передачи данных по каналам между устройствами хранения и обработки, а также их обработки на соответствующих вычислительных устройствах. Особенностью математической модели процесса распределенного хранения и обработки данных, модели иерархической игры оптимизации решений является учет длительностей интервалов времени их (данных) пребывания на устройствах хранения. Эта характеристика процесса определяемых в соответствии с сформированными на нижнем уровне расписаниями обработки, а также длительностями интервалов времени выставления данных в каналы и их передачи по каналам.

В качестве способа оптимизации решений по распределенному хранению и обработке данных на каждом из уровней иерархии использованы генетические алгоритмы. Выполнена разработка способов кодирования решений на каждом из уровней, а также генетических операторов, позволяющих осуществлять поиск локально оптимальных решений.

С целью построения расписаний обработки данных, закрепленных за соответствующими вычислительными устройствами, разработан эвристический алгоритм, предусматривающий упорядочивание обработки данных с точки зрения не убывания их объема. Использование разработанного эвристического алгоритма построения расписаний обработки данных на устройствах обеспечивает минимизацию интервалов времени их нахождения на устройствах хранения и, как следствие, минимизацию затрат на хранение. Выполненные исследования эффективности применения разработанного метода многоуровневой оптимизации решений по распределенным хранению и обработки данных показали, что его использование позволяет на 10–60 % снизить суммарные финансовые затраты на выполнения указанных операций, а также на операции передачи данных и штрафы за неполное использование ограниченных ресурсов.

Практическая ценность полученных результатов состоит в разработке алгоритмов оптимизации решений по распределенному хранению и обработке данных, которые могут быть непосредственно применены при планировании вычислений в кластерных и облачных системах.

Список источников

1. Prajapati H.B., Shah V.A. Scheduling in Grid Computing Environment // Proceedings of the Fourth International Conference on Advanced Computing & Communication Technologies (Rohtak, India, 08–09 February 2014). IEEE, 2014. PP. 315–324. DOI:10.1109/ACCT.2014.32

2. Bhatia M.K. Task Scheduling in Grid Computing: A Review // Advances in Computational Sciences and Technology. 2017. Vol. 10. Iss. 6. PP. 1707–1714.

3. Xhafa F., Barolli L., DurresiA. Batch mode scheduling in grid systems // International Journal of Web and Grid Services. 2007. Vol. 3. Iss. 1. PP. 19–37. DOI:10.1504/IJWGS.2007.012635

4. Khan M. Design and Analysis of Security Aware Scheduling in Grid Computing Environment // International Journal of Computer Science and Information Technology Research (IJCSITR). 2013. Vol. 1.Iss. 1. PP. 42–50.

5. Naresh U. Study on Many-Task-Computing using Data Aware Scheduling in Cloud Computing // International Journal of Innovations & Advancement in Computer Science (IJIACS). 2017. Vol. 6. Iss. 9. PP. 360–366.

6. Mahajan S., Kaur R. A Concern towards Job scheduling in Cluster Computing // International Journal of Computer Engineering in Research Trends. 2015. Vol. 2. Iss. 6. PP. 392–394.

7. Abawajy J.H. Dynamic Parallel Job Scheduling in Multi-cluster Computing Systems // Proceedings of the 4th International Conference of Computer Science (ICCS 2004, Kraków, Poland, 6–9 June 2004). Lecture Notes in Computer Science. Vol. 3036. Berlin, Heidelberg: Springer, 2004. PP. 27–34. DOI:10.1007/978-3-540-24685-5_4

8. Alworafi M.A., Dhari A., El-Booz Sh.A., Mallappa S. Budget-aware task scheduling technique for efficient management of cloud resources // International Journal High Performance Computing and Networking. 2019. Vol. 14. Iss. 4. PP. 453–465. DOI:10.1504/IJHPCN.2019.102352

9. Arabnejad V., Bubendorfer K., Ng B. Budget and Deadline Aware e-Science Workflow Scheduling in Clouds // IEEE Transactions on Parallel and Distributed Systems. 2019. Vol. 30. Iss. 1. PP. 29–44. DOI:10.1109/TPDS.2018.2849396

10. Месарович М., Мако Д., Такахара И. Теория иерархических многоуровневых систем. М.: Из-во «Мир», 1973. 344 с.

11. Воронин А.А., Мишин С.П. Оптимальные иерархические структуры. М.: ИПУ РАН, 2003. 214 с.

12. Губко М.В., Новиков Д.А. Теория игр в управлении организационными системами. М.: Институт проблем управления им. В.А. Трапезникова, 2005. 138 с.

13. Бурков В.Н., Коргин Н.А., Новиков Д.А. Введение в теорию управления организационными системами. М.: Либроком, 2009. 264 с.
14. Бусыгин В.П., Желободько Е.В., Коковин С. Г., Цыплаков А.А. Микроэкономический анализ несовершенных рынков. Новосибирск: Новосиб. гос. ун-т, 1999. 132 с.
15. Гладков Л.А., Курейчик В.В., Курейчик В.М. Генетические алгоритмы. М.: Физматлит, 2006. 320 с.
16. Курейчик В.М. Генетические алгоритмы и их применение. Таганрог: Таганрогское РТУ, 2002. 244 с.
17. Смирнов А.В. О задаче упаковки в контейнеры // Успехи математических наук.1991. Т. 46. № 4. С. 173–174.

References

1. Prajapati H.B., Shah V.A. Scheduling in Grid Computing Environment. *Proceedings of the Fourth International Conference on Advanced Computing & Communication Technologie, 08–09 February 2014s, Rohtak, India*. IEEE; 2014. p.315–324. DOI:10.1109/ACCT.2014.32
2. Bhatia M.K. Task Scheduling in Grid Computing: A Review. *Advances in Computational Sciences and Technology*. 2017;10(6):1707–1714.
3. Xhafa F., Barolli L., Durresi A. Batch mode scheduling in grid systems. *International Journal of Web and Grid Services*. 2007;3(1):19–37. DOI:10.1504/IJWGS.2007.012635
4. Khan M. Design and Analysis of Security Aware Scheduling in Grid Computing Environment. *International Journal of Computer Science and Information Technology Research (IJCSITR)*. 2013;1(1):42–50.
5. Naresh U. Study on Many-Task-Computing using Data Aware Scheduling in Cloud Computing. *International Journal of Innovations & Advancement in Computer Science (IJIACS).* 2017;6(9):360–366.
6. Mahajan S., Kaur R. A Concern towards Job scheduling in Cluster Computing. *International Journal of Computer Engineering in Research Trends.* 2015;2(6):392–394.
7. Abawajy J.H. Dynamic Parallel Job Scheduling in Multi-cluster Computing Systems. *Proceedings of the 4th International Conference of Computer Science, ICCS 2004, 6–9 June 2004, Kraków, Poland. Lecture Notes in Computer Science.* Berlin, Heidelberg: Springer; 2004. *vol.3036*. p. 27–34. DOI:10.1007/978-3-540-24685-5_4
8. Alworafi M.A., Dhari A., El-Booz Sh.A., Mallappa S. Budget-aware task scheduling technique for efficient management of cloud resources. *International Journal High Performance Computing and Networking*. 2019;14(4):453–465. DOI:10.1504/IJHPCN.2019.102352
9. Arabnejad V., Bubendorfer K., Ng B. Budget and Deadline Aware e-Science Workflow Scheduling in Clouds. *IEEE Transactions on Parallel and Distributed Systems*. 2019;30(1):29–44. DOI:10.1109/TPDS.2018.2849396
10. Mesarovich M., Mako D., Takahara I. *Theory of Hierarchical Multilevel Systems*. Moscow: Mir Publ.; 1973. 344 p. (in Russ.)
11. Voronin A.A., Mishin S.P. *Optimal Hierarchical Structures*. Moscow: V.A. Trapeznikov Institute of Management Problems Publ.; 2003. 214 p. (in Russ.)
12. Gubko M.V., Novikov D.A. *Game Theory in the Management of Organizational Systems.* Moscow: V.A. Trapeznikov Institute of Management Problems Publ.; 2005. 138 p. (in Russ.)
13. Burkov V.N., Korgin N.A., Novikov D.A. *Introduction to the Theory of Management of Organizational Systems*. Moscow: Librocom Publ.; 2009. 264 p.(in Russ.)
14. Busygin V.P., Zhelobodko E.V., Kokovin S.G., Tsyplakov A.A. *Microeconomic Analysis of Imperfect Markets*. Novosibirsk: Novosibirsk State University Publ.; 1999. 132 p. (in Russ.)
15. Gladkov L.A., Kureychik V.V., Kureychik V.M. *Genetic Algorithms*. M.: Fizmatlit Publ.; 2006. 320 p. (in Russ.)
16. Kureychik V.M. *Genetic Algorithms and Their Application*. Taganrog: Taganrog Radio Engineering University Publ.; 2002. 244 p. (in Russ.)
17. Smirnov A.V. On the Problem of Packaging in Containers. *Successes of Mathematical Sciences.* 1991;46(4):173–174. (in Russ.)

Статья поступила в редакцию 28.11.2022; одобрена после рецензирования 23.01.2023; принята к публикации 26.01.2023.
The article was submitted 28.11.2022; approved after reviewing 23.01.2023; accepted for publication 26.01.2023.

Информация об авторе:

КРОТОВ Кирилл Викторович
доктор технических наук, доцент, доцент кафедры «Информационные системы» Севастопольского государственного университета
https://orcid.org/0000-0002-9670-6141

Научная статья
УДК 004.056(075.8)
DOI:10.31854/1813-324X-2023-9-2-128-142

Методы защиты от угрозы неправильного заполнения избирательного бюллетеня в системе дистанционного электронного голосования

Виктор Алексеевич Яковлев, yakovlev.va@sut.ru
Васан Давуд Салман, salman.vd@sut.ru

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

Аннотация: *Рассматривается обобщенная схема дистанционного электронного голосования, основанная на гомоморфном шифровании. Исследуются два метода защиты системы голосования от угрозы со стороны избирателя, заключающиеся в неправильном заполнении бюллетеня избирателем. Оба метода основаны на алгоритмах «доказательства с нулевым разглашением секрета». Получены оценки сложности вычислений при формировании доказательства корректности заполнения бюллетеня избирателем и оценки сложности проверки доказательства контролирующей стороной. Сравнительный анализ сложности реализации обоих методов показал, что метод, основанный на доказательстве на базе равенства логарифмов) имеет меньшую сложность вычислений на стороне избирателя по сравнению с методом, основанном на перемешивании голосов избирателей. В тоже время второй метод (перемешивания голосов) требует в 1,67 раза меньше вычислений в блокчейне, что становится существенным фактором выбора в пользу второго метода при большом количестве избирателей.*

Ключевые слова: *система дистанционного электронного голосования, схема Эль-Гамаля на эллиптической кривой, схема перемешивания, проверка доказательства корректности заполнения бюллетеня, доказательство с нулевым разглашением секрета*

Ссылка для цитирования: Яковлев В.А., Салман В.Д. Методы защиты от угрозы неправильного заполнения избирательного бюллетеня в системе дистанционного электронного голосования // Труды учебных заведений связи. 2023. Т. 9. № 2. С. 128–142. DOI:10.31854/1813-324X-2023-9-2-128-142

Methods of Protection against Threat: Incorrect Ballot Filling by Voter in the Remote Electronic Voting System

Victor Yakovlev, yakovlev.va@sut.ru
Vasan Salman, salman.vd@sut.ru

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

Abstract: *A generalized scheme of remote electronic voice based on homomorphic encryption is considered. Two methods of protecting the voting system from the threat from the voter, consisting in incorrect filling of the ballot by*

the voter, are investigated. Both methods are based on the algorithms of "zero-knowledge proof". Evaluations of the complexity of calculations in the formation of proof of the correctness of filling in the ballot by the voter and Evaluations of the complexity of verification of the proof by the controlling party are obtained. A comparative analysis of the complexity of the implementation of both methods has shown that the method based on the proof based on the equality of logarithms has less complexity of calculations on the voter's side compared to the method based on the mixing of votes. At the same time, the second method (the method of mixing votes) requires 1.67 times less calculations in the blockchain, which becomes a significant factor in choosing the second method in favor of a large number of voters.

Keywords: *the elliptic ElGamal scheme, remote electronic voting system, mixing scheme, verification scheme, proof of filling the ballot, zero-knowledge proof system*

For citation: Yakovlev V., Salman V. Methods of Protection against Threat: Incorrect Ballot Filling by Voter in the Remote Electronic Voting System. *Proc. of Telecom. Universities.* 2023;9(2):128–142. (in Russ.) DOI:10.31854/1813-324X-2023-9-2-128-142

Введение

Системы дистанционного электронного голосования (ДЭГ) все шире входят в жизнь современного общества. Распространение получили системы электронного голосования на основе микс-сетей [1–4], на основе слепой подписи [5–7] и на основе гомоморфного шифрования [8–10]. В этих системах решаются две главные задачи: обеспечение тайны и анонимности голосования, в том числе и для избирательной комиссии.

Принцип работы на основе микс-сетей заключается в создании системы из нескольких связанных прокси-серверов, которые называют миксами. Клиент шифрует сообщение один раз с использованием открытых ключей каждого из прокси-серверов в определенном порядке, который знает только он. Расшифровка криптограммы происходит в обратном порядке с помощью секретных ключей микс-серверов, но уже на стороне последних. Так как голоса приходят в избирательную комиссию в «перепутанном» виде, обеспечивается анонимность голосования. Система электронного голосования, основанная на слепой подписи, представляет собой криптографический метод, в котором сообщение m избирателя A подписывается органом подписи B, так что B не получает никакой информации о сообщении m. При этом обеспечивается доверие к переданному сообщению, но сохраняется анонимность избирателя. В системе голосования на основе гомоморфных криптосистем последние зашифровывают свои бюллетени открытым ключом избирательной комиссии. Затем они отправляют свои зашифрованные бюллетени на сервер, который «перемножает» все бюллетени и отправляет получивший результат в избирательную комиссию. Та расшифровывает это произведение бюллетеней и объявляет победителя выборов. Так как расшифрование выполняется сразу всех агрегированных бюллетеней, то обеспечивается анонимность каждого избирателя.

Известны практические системы голосования, в разной степени использующие эти подходы.

Во-первых, ДЭГ в России (см. URL: https://evoting.digitaldem.ru/wp-content/uploads/sites/2/2021/07/ptkdeg_general_description_2021-07-15.pdf). Организатор голосования (Комиссия ДЭГ) и Учетчик (блокчейн) генерируют ключевые пары (ключи шифрования и расшифрования бюллетеней). На блокчейне (БЧ) формируется итоговый открытый ключ шифрования, который передается Регистратору и избирателю. Закрытый ключ разделяется на доли. Избиратель генерирует ключевую пару электронной подписи. Избиратель и Регистратор выполняют протокол формирования подписи вслепую для ключа проверки электронной подписи избирателя. Избиратель заполняет бюллетень из значений 1 – «за» и 0 – «против», шифрует их с помощью ключа шифрования бюллетеней, формирует доказательство корректности содержимого бюллетеня, состоящее в том, что его выбор соответствует либо 0, либо 1. Также формируется доказательство корректности заполнения бюллетеня в целом.

Во-вторых, Provotum (Швейцария) [8]. Каждый сервер генерирует свой собственный открытый ключ (pk); на БЧ формируются общий открытый ключ (pk_{voting}). Далее избиратель заполняет бюллетень из значений, шифрует их с помощью общего ключа и формирует доказательство корректности содержимого бюллетеня, состоящее в том, что его выбор соответствует либо 0, либо 1. Система отличается от других тем, что каждый сервер генерирует свой собственный открытый ключ, а общий открытый ключ формируется в БЧ.

В-третьих, Helios (США) [4, 12]. Сначала сервер генерирует бланк – бюллетень; далее избиратель выбирает своего кандидата из значений (0, 1), и сервер шифрует выбор избирателя, используя открытый ключ; после чего отправляет все зашифрованные бюллетени к микс-серверу, который маскирует и перемешивает их. Микс-сервер также должен доказать, что правильно перемешал бюллетени.

Примечание. Во всех рассмотренных системах использована схема Эль-Гамаля [11] на эллиптической кривой для шифрования и расшифрования бюллетеней.

Для всех систем голосования существует достаточно много угроз, связанных с действиями нарушителя и неправомерными действиями участников протокола голосования [13–15]. В последнее время большое внимание при построении систем электронного голосования уделяется защите от угрозы преднамеренного или непреднамеренного неправильного заполнения бюллетеня голосования избирателем. Эта задача не является тривиальной, так как контроль правильности заполнения бюллетеня должен осуществляться в зашифрованном виде, без раскрытия того, как проголосовал избиратель.

В [16–18] рассматривается протокол электронного голосования с проверкой корректности заполнения бюллетеней. Протокол работает следующим образом: сначала избиратель шифрует свой бюллетень и получает криптограмму B_i. Далее он должен доказать, что в криптограмме зашифрованы значения $(0, 1)$. Для этого формируется доказательство корректности заполнения своего бюллетеня. Криптограмма и доказательство отправляется в избирательную комиссию, которая проверяет доказательства для (B_i): если проверка прошла успешно, то голос избирателя принимается. Далее комиссия расшифровывает и подсчитывает голоса.

В [19] предложена система ДЭГ, использующая гомоморфную схему. В этой работе доказательство корректности заполнения бюллетеня ИК и его проверка разработаны для общего случая, когда вариант выбора избирателя принадлежит заданному диапазону возможных значений. Сложность такого доказательства в значительной степени зависит от количества возможных вариантов голосования на выборах.

В [20] предложена система ДЭГ на основе гомоморфного шифрования, в которой для доказательства корректности заполнения бюллетеня использована схема перемешивания голосов, поданных за кандидатов. Эта схема основывается на работах [2, 3, 21], в которых представлены доказательства корректности выполнения этой процедуры.

Все проверки корректности заполнения бюллетеня выполняются с использованием неинтерактивных схем доказательств с нулевым разглашением секрета.

Целью работы является исследование методов защиты от угрозы неправильного заполнения бюллетеня избирателем в системе ДЭГ, оценка сложности их реализаций и рекомендации по их применению. В п. 1 приведена модель системы ДЭГ, на основе схемы шифрования Эль-Гамаля на эллиптической кривой и угроз, специфических для системы. В п. 2 приведено детальное описание метода проверки корректности заполнения бюллетеня, основанного на доказательстве с нулевым разглашением секрета в задаче дискретного логарифмирования. В п. 3 приведено описание метода проверки корректности заполнения бюллетеня на основе перемешивания голосов избирателя. Описание методов сопровождается числовыми примерами правильного и неправильного заполнения бюллетеня. В п. 4 проведен сравнительный анализ сложности реализации обоих методов и рекомендации по их использованию в системах ДЭГ.

1. Модель системы ДЭГ на основе схемы шифрования Эль-Гамаля на эллиптической кривой

Рассматриваемая в работе система ДЭГ включает в себя: избирателей, сервер, БЧ и ИК (рисунок 1).

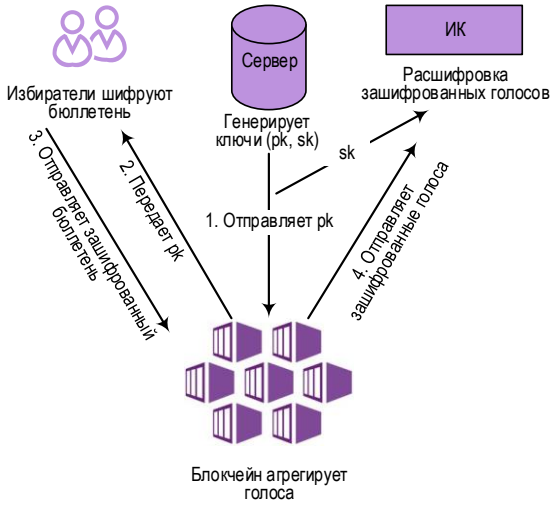


Рис. 1. Схема ДЭГ

Fig. 1. Remote Electronic Voting Scheme

Рассмотрим систему ДЭГ, построенную на основе гомоморфной системы шифрования Эль-Гамаля [11]. Под гомоморфным шифрованием понимается криптографический примитив, представляющий собой функцию шифрования, удовлетворяющую дополнительному требованию гомоморфности относительно каких-либо алгебраических операций над открытыми сообщениями. Свойство гомоморфного шифрования позволяет агрегировать голоса в зашифрованном виде и после расшифровать одну криптограмму, получив сразу результат голосования.

Основными этапами функционирования системы являются:

– инициализация системы;
– аутентификация избирателей;
– голосование и подсчет голосов;
– объявление результатов голосования.

Инициализация системы заключается в выборе системных параметров и генерации ключей. Сервер генерирует открытый и закрытый ключ для криптосистемы, использующей гомоморфное шифрование, и отправляет открытый ключ в БЧ, который передает открытый ключ всем избирателям. Сек-

ретный ключ хранится на сервере или может быть разделен на доли и находиться у хранителей ключа до окончания выборов. После того, как избиратель успешно пройдет этап идентификации и аутентификации, он получает разрешение на участие в голосовании (в работе процесс аутентификации и идентификации избирателя не рассматривается).

Каждый избиратель выбирает кандидата/кандидатов из списка, шифрует свой голос с помощью открытого ключа и отправляет его в БЧ. После завершения голосования в БЧ осуществляется агрегирование голосов, результаты отправляются в избирательную комиссию. Сервер, на котором генерировались открытый и закрытый ключи, передает закрытый ключ избирательной комиссии, а если было разделение ключа, доверенные лица передают свои доли ключа, комиссия, в свою очередь, восстанавливает закрытый ключ. Далее она расшифровывает результаты голосования с помощью закрытого ключа и объявляет итог.

Одна из угроз в данной системе ДЭГ заключается в том, что избиратель может неправильно (умышленно или случайно) заполнить свой бюллетень, и это повлияет на результаты голосования. Чтобы предотвратить эту угрозу, применяются различные методы проверки корректности заполнения бюллетеня. В работе проведен сравнительный анализ двух методов решения этой задачи: основанного на сравнении дискретных логарифмов [16–19] и на проверке корректности перестановки [2, 3, 20]. Оба метода относятся к задачам «доказательства с нулевым разглашением секрета» [22–26].

Рассмотрим далее модель системы ДЭГ на основе схемы гомоморфного шифрования Эль-Гамаля на эллиптической кривой [27–28]. Эта схема и параметры кривой будут далее использоваться для шифрования бюллетеня и выполнения других функций во всей работе.

Генерация ключей

Сервер генерирует эллиптическую кривую вида $y^2 = x^3 + xa + b$ над полем Галуа $GF(p)$ и выбирает базовую точку $P \in E\big(GF(p)\big)$ порядка m.

Сервер случайным образом выбирает закрытый ключ d, $d \in \{1, \dots, m-1\}$. Далее вычисляется открытый ключ: $Q = dP \bmod p$ и генерируется точка $F = rP \bmod$, где r – случайное число, выбираемое в диапазоне $[1, \dots, m-1]$.

Параметры p, E, m, P, F, Q публикуются в БЧ. Секретный ключ d хранится в избирательной комиссии в разделенном на доли виде.

Шифрование бюллетеня

Избиратель V_i, i = 1, 2, …, n, где n – количество избирателей, шифрует сообщение (бюллетень) M_i по схеме Эль-Гамаля с помощью открытого ключа и получает криптограмму:

$$\mathrm{Enc}(M_i) = C_i = (A_i, B_i), \quad (1)$$

где Enc() – функция шифрования; (A_i, B_i) – две части криптограммы C_i,: первая часть $A_i = rP \bmod p$; вторая часть $B_i = (M_i F + rQ) \bmod p$; r – выбирается случайным образом.

Дешифрование бюллетеня

Расшифрование криптограммы осуществляется с помощью закрытого ключа d:

$$\mathrm{Dec}(C_i) = B_i - dA_i \bmod p, \quad (2)$$

где Dec() – функция дешифрования.

Результат расшифровки должен быть равен сообщению M_i.

Криптосистема Эль-Гамаля на эллиптической кривой обладает гомоморфным свойством.

Допустим, есть два шифртекста:

$$C_1 = (A_1, B_1) = (r_1 P, F_1 + r_1 Q) \text{ и} \quad (3)$$

$$C_2 = (A_2, B_2) = (r_2 P, F_2 + r_2 Q). \quad (4)$$

Криптограммы могут быть агрегированы аддитивно:

$$C_3 = C_1 + C_2 = \\ = \big((r_1 + r_2)P,\ (F_1 + F_2) + (r_1 + r_2)Q\big). \quad (5)$$

Тогда при расшифровании C_3 получаем:

$$\mathrm{Dec}(C_3) = F_1 + F_2. \quad (6)$$

Рассмотрим далее способ заполнения бюллетеня.

Заполнение бюллетеня

Бюллетень в электронном виде представляет собой строку символов (1, 0). В зависимости от правил выборов могут быть различные варианты голосования. Например, избиратель может проголосовать за одного кандидата из k кандидатов, или он может проголосовать за двух и более кандидатов (t из N). Но он не может не голосовать. Могут быть и другие правила, установленные избирательной комиссией. Любые отклонения от установленных вариантов голосования, например, использование числа 2 или –1, поданных за какого-то кандидата, будут означать некорректное заполнение бюллетеня. Пример правильного заполнения бюллетеня показан в таблице 1. Избиратель подал голос «за» за первого и четвертого кандидатов, и голос «против» – за остальных кандидатов. Таким образом, бюллетень должен содержать только значения (1, 0). Для того, чтобы подтвердить, что он действительно заполнил свой бюллетень правильно, необходимо использовать методы доказательства корректности заполнения бюллетеня.

ТАБЛИЦА 1. Формирование правильного заполнения бюллетеня

TABLE 1. Formation of the Correct Filling of the Ballot

Кандидаты	*D*1	*D*2	*D*3	*D*4	……	*Dk*
Выбор избирателя	1	0	0	1	….....	0

2. Метод проверки корректности заполнения бюллетеня на основе проверки логарифмов

2.1. Проверка корректности заполнения бюллетеня для каждого шифртекста

Рассмотрим [19] протокол голосования, когда выбирается только один кандидат из k кандидатов. Избиратель может голосовать («за» одного и «против» остальных кандидатов). Проверка доказательств осуществляется на основе неинтерактивного метода с нулевым разглашением секрета (NIZKP, *аббр. от англ.* Non-Interactive Zero-Knowledge Proof) и заключатся в доказательстве сравнения вида:

$$ZP(x|y(x) = z), \quad (7)$$

где x – параметр, не неизвестный проверяющему; z – известная проверяющему величина.

В нашем случае нужно доказать, что для каждой криптограммы C_i выполняется сравнение:

$$ZP(r_i, b_{vi}|C_i = (r_iP, b_{vi}F_i + r_iQ) \vee (C_i = (r_iP, r_iQ)),$$

где b_{vi} – голос i-го избирателя, $b_{vi} \in \{0,1\}$; r_i – случайное число.

Алгоритм голосования, формирование доказательства корректности заполнения бюллетеня и проверки доказательства для вышерассмотренной схемы голосования включает следующие шаги и приведен в таблице 2.

Шаг 1. Загрузка открытого ключа из БЧ.

Шаг 2. Выбор своего кандидата.

Шаг 3. Шифрование бюллетеня по схеме Эль-Гамаля на эллиптической кривой.

Шаг 4. Формирование доказательства того, что он зашифровал свой бюллетень из значений (1, 0).

Последняя колонка (см. таблицы 2 и 3) содержит оценки сложности выполнения соответствующих операций. Символ M обозначает операцию умножения точки эллиптической кривой на целое число. Операции сложения точек не учитывались ввиду их меньшей сложности по сравнению с операцией умножения; H – сложность операции хеширования также не учитывалась.

Далее избиратель отправляет значения (A, B, a_1, b_1, a_2, b_2, u_1, u_2, t_1, t_2) проверяющему (в БЧ), где, согласно алгоритму из таблицы 3, проходит проверка того, что избиратель правильно заполнил свой бюллетень. Здесь же приведены оценки сложности выполнения алгоритма. Если все сравнения выполняются, значит избиратель правильно проголосовал за каждого кандидата при этом проверяющий (БЧ) не знает, как проголосовал избиратель.

ТАБЛИЦА 2. Формирование доказательства корректности заполнения бюллетеня

TABLE 2. Formation of Proof of Correctness of Filling in the Ballot

Избиратель: голосование и формирование доказательства			Оценки сложности (при выборе b_v = 1)
Голосует:	«за» кандидата – $b_{vi} = 1$	«против» кандидата – $b_{vi} = 0$	$O(1)$
Случайным образом выбирает числа $w, r_1, t_1, u_1 \in Z_q$.			$O(1)$
Осуществляет шифрование бюллетеня по каждому кандидату (вычисляет):	$A = (r_1P) \bmod p$; $B = (b_{vi}F + r_1Q) \bmod p$.	$A = (r_1P) \bmod p$; $B = (r_1Q) \bmod p$.	1kM 2kM
Формирует доказательство корректности голосования (вычисляет):	$a_1 = (t_1P - u_1A) \bmod p$; $b_1 = (t_1Q - u_1(B - b_{vi}\mathrm{P}) \bmod p$ $a_2 = wP \bmod p$; $b_2 = wQ \bmod p$.	$a_1 = \mathrm{wP} \bmod p$; $b_1 = wQ \bmod p$; $a_2 = (t_1\mathrm{P} - u_2\mathrm{A}) \bmod p$; $b_2 = (t_1\mathrm{Q} - u_2(\mathrm{B} - b_{vi}\mathrm{P}) \bmod p$	2kM 3kM 1kM 1kM
Вычисляет хэш-функцию $h = H(A, B, a_1, b_1, a_2, b_2) \bmod q$			1H
Вычисляет доказательство:	$h - u_1 \bmod q$; $t_2 = w - r_1u_2 \bmod q$.	$u_1 = h - u_2 \bmod q$; $t_1 = w - r_1u_1 \bmod q$.	$O(k)$ $O(k)$
Всего операций умножения точки эллиптической кривой на число			10kM

ТАБЛИЦА 3. Алгоритм проверки корректности голосования за кандидата

TABLE 3. Algorithm for Verifying the Correctness of Voting for a Candidate

Проверяющий (БЧ)			Оценки сложности
Вычисляет хэш-функцию $h = H(A, B, a_1, b_1, a_2, b_2)$			1H
Проверяет сравнения:	$h \bmod q =^? u_1 + u_2 \bmod q$; $t_1P \bmod p =^? a_1 + u_1A \bmod p$; $t_1Q \bmod p =^? b_1 + u_1(B - b_iP) \bmod p$.	(8) (9) (10)	$O(k)$ 2kM 3kM
Всего операций умножения точки эллиптической кривой на число			5kM

Примечание. В таблицах приняты следующие условные обозначения: H – сложность операции хеширования; M – операция умножения точки эллиптической кривой на целое число; k – количество кандидатов.

Рассмотрим примеры формирования и проверки доказательства корректности заполнения бюллетеня для варианта, когда выбирается один кандидат $D1$ из 4 кандидатов. Пусть на этапе инициализации системы ДЭГ выбраны параметры: $p = 59$, $q = 17$, $(a = 3, b = 9)$. Эллиптическая кривая является несингулярной и имеет следующие точки (выбор кривой и параметров шифрования носят иллюстрационный характер):

{(0, 3), (0, 56), (3, 24), (3, 35), (4, 12), (4, 47), (6, 19), (6, 40), (7, 14), (7, 45), (9, 23), (9, 36), (10, 6), (10, 53), (11, 4), (11, 55), (12, 11), (12, 48), (13, 11), (13, 48), (14, 9), (14, 50), (15, 19), (15, 40), (17, 28), (17, 31), (19, 9), (19, 50), (20, 24), (20, 35), (25, 29), (25, 30), (26, 9), (26, 50), (29, 0), (34, 11), (34, 48), (36, 24), (36, 35), (38, 19), (38, 40), (42, 1), (42, 58), (46, 29), (46, 30), (47, 29), (47, 30), (49, 10), (49, 49), (50, 16), (50, 43), (51, 2), (51, 57), (54, 20), (54, 39), (55, 13), (55, 46), (58, 8), (58, 51), $(\mathcal{O}, \mathcal{O})$}.

Используя схему Эль-Гамаля и выбрав базовую точку $P = (19, 9)$, генерируются ключи: закрытый – $d = 4$ и открытый – $Q = 4(19, 9) \bmod 59 = (6, 19)$. Далее выбирается случайным образом $r = 3$ и вычисляется $F = rP \bmod p = 3(19, 9) \bmod 59 = (54, 39)$. Параметры p, E, m, P, F, Q публикуются в БЧ. (Порядок выполнения операций сложения и умножения точек эллиптической кривой на целое число можно найти в [29]).

Примечание. В дальнейшем будем предполагать, что для вычисления хэш-функции используется некоторый алгоритм, вырабатывающий по заданному аргументу число, которое мы в числовых примерах указываем произвольно.

Рассмотрим два «полярных» случая.

Случай 1. Избиратель правильно заполнил свой бюллетень:

– избиратель V_1 голосует «за» $(b_{v1} = 1)$, выбирает случайным образом $r_1 = 2$;

– шифрует свой бюллетень:

$$(A_1, B_1) = (r_1 P, b_{v1} F + r_1 Q) \bmod p;$$

$\big(2(19, 9) + 1(52, 2) + 2(6, 19)\big) \bmod 59 =$ ((49, 10), (6, 40));

– вычисляет доказательство для криптограммы $(A_1, B_1) = ((49, 10), (6, 40))$; если $b_{v1} = 1$, выполняет вычисления согласно второму столбцу таблицы 2:

а) случайным образом выбирает числа: $t_1 = 2$, w $= 2, u_1 = 5$;

б) вычисляет:

$$a_1 = \big(2(19, 9) - 5(49, 10)\big) \bmod 59 = (34, 48);$$

$$b_1 = (2(6, 19) - 5\big((6, 40) - 1(19, 9)\big) \bmod 59 = (54, 39);$$

$$a_2 = 2(19, 9) \bmod 59 = (49, 10);$$

$$b_2 = 2(6, 19) \bmod 59 = (34, 11).$$

Предположим, что хеширование параметров $(A_1, B_1, a_1, b_1, a_2, b_2)$ дает $h = 3$:

– вычисляет:

$u_2 = 3 - 5 \bmod 17 = 15; t_2 = 2 - 2 \times 15 \bmod 17 = 6;$

– отправляет в БЧ зашифрованный бюллетень:

$$(A_1 = (49, 10), B_1 = (6, 40))$$

и доказательство:

$$\begin{pmatrix} a_1 = (34, 48), b_1 = (54, 39), a_2 = (49, 10), \\ b_2 = (34, 11), u_1 = 5, u_2 = 15, t_1 = 2, t_2 = 6 \end{pmatrix}.$$

Таким же образом шифруются голоса для остальных кандидатов:

$$C_2 = ((6, 19), (19, 9)), C_3 = ((51, 2), (51, 2)),$$

$$C_4 = ((34, 48), (49, 49)),$$

где криптограммы C_2, C_3, C_4 являются зашифрованными значениями точки $\mathcal{O}$ и вычисляются доказательства для этих криптограмм согласно третьему столбцу таблицы 2.

БЧ проверяет, что избиратель правильно заполнил свой бюллетень, выполняя сравнение согласно таблице 3.

Для нашего примера – БЧ:

– вычисляет хэш-функцию $h = 3$, находит:

$$(u_1 + u_2) \bmod q = (5 + 15) \bmod 17 = 3$$

(видим, что сравнение (8) выполняется: $3 = 3$);

– находит $t_1 P \bmod p = (49, 10)$ и $a_1 + u_1 A_1 \bmod p = (49, 10)$;

– проверяет, что сравнение (9) выполняется:

$$t_1 P \bmod p \stackrel{?}{=} a_1 + u_1 A_1 \bmod p; (49, 10) = (49, 10).$$

– вычисляет:

$$t_1 \mathrm{Q} \bmod p = (34, 11)$$

и

$$b_1 + u_1 (B_1 - b_{vi} P) \bmod p = (34, 11);$$

(видим, что сравнение (10) выполняется: $(34, 11) = (34, 11)$.

Таким образом, все сравнения выполнены, следовательно, корректность заполнения бюллетеня для $D1$ доказана.

Аналогично проверяются доказательства корректности голосования за других кандидатов.

Случай 2. Избиратель неправильно заполнил свой бюллетень.

Пусть избиратель поставил число 2 за $D1$. Все шаги алгоритма аналогичны предыдущему примеру:

– при выборе в $b_{v1} = 2$ находит:

$$(A_1, B_1) = (r_1 P, b_{v1} F + r_1 Q) \bmod p =$$

$$= \big((54, 39), (54, 39)\big);$$

– формирует доказательство:

$$a_1 = (6, 40), b_1 = (11, 4), a_2 = (49, 10), b_2 = (34, 11),$$

$$h = 5, u_2 = 3, t_2 = 13.$$

БЧ проверяет доказательство корректности заполнения бюллетеня, проверяя сравнения согласно таблице 3:

– сравнение (8) выполняется: $h \bmod q \stackrel{?}{=} u_1 + u_2 \bmod q; 5 = 5;$

– сравнение (9) выполняется: $(49, 10) = (49, 10);$

– сравнение (10) не выполняется: $(34, 11) \neq \neq (49, 10)$.

Видно, что не все сравнения выполнены, следовательно, корректность заполнения бюллетеня для $D1$ не доказана.

2.2. Проверка корректности заполнения всего бюллетеня

В случае, рассмотренном выше, контролирующий орган может убедиться, что избиратель корректно проголосовал за каждого кандидата («за» или «против»). Но он не может проверить, выполнены ли правила голосования по заданному варианту голосования. То есть, например, избиратель может выбрать трех кандидатов, хотя разрешено выбрать только одного или двух. Эта задача решается проверкой корректности заполнения бюллетеня в целом (см. URL: https://evoting.digitaldem.ru/wp-content/uploads/sites/2/2021/07/ptkdeg_general_description_2021-07-15.pdf). Рассмотрим этот метод.

Пусть $k_{\max}$ – максимальное число голосов «за», при голосовании за k кандидатов. Будем считать, что ключи (открытый, закрытый) сгенерированы, избиратель выполнил следующие действия:

– выбрал кандидатов;

– зашифровал бюллетень с помощью открытого ключа: $C_i = (A_i, B_i) \bmod p$, где $A_i = r_i P \bmod p$; $B_i = F + r_i Q \bmod p$, если $F = b_{vi} P \bmod p$, b_{vi} – выбор избирателем кандидата, $b_{vi} \in \{0,1\}, i = 1, 2, \ldots, k$.

– сформировал доказательство корректности голосования за каждого кандидата, как было описано выше.

Рассмотрим подробно формирование доказательства корректности заполнения бюллетень.

Избиратель вычисляет сумму криптограмм бюллетеня для всех кандидатов:

$$C_\Sigma = (A_\Sigma, B_\Sigma), \tag{11}$$

где $A_\Sigma = \sum_{i=1}^{k} A_i$, $B_\Sigma = \sum_{i=1}^{k} B_i$, $r = \sum r_i, m = \sum m_i$, m – сумма голосов «за», поданных избирателем в пользу всех кандидатов.

Выполняет следующий алгоритм:

1) находит:

$$T = t \cdot Q, \tag{12}$$

где $t \in Z_p$ – случайное число;

2) вычисляет хэш-функцию:

$$h = H(Q, A_\Sigma, B_\Sigma, T, m); \tag{13}$$

3) вычисляет:

$$s = t + r \cdot h; \tag{14}$$

4) посылает в БЧ (T, s, m').

Избиратель с целью обмана может указать суммарное число голосов «за», поданных в пользу всех кандидатов m', отличное от фактического числа голосов m, если $m > k_{\max}$.

БЧ вычисляет: $h = H(Q, \sum A_i, \sum B_i, T, m')$, для чего используются криптограммы $C_i = (A_i, B_i)$ из бюллетеня.

Далее БЧ проверяет сравнение:

$$sQ \stackrel{?}{=} T + h\left(\sum_{i=1}^{k} B_i - m' F_i\right). \tag{15}$$

Если сравнение выполнятся, то $m = m'$. Покажем, что это действительно так:

$$T + h\left(\sum_{i=1}^{k} B_i - m' F_i\right) = tQ + h\left(mF + r_\Sigma Q - m' F\right) =$$

$$= tQ + r_\Sigma \mathrm{h} Q + h(mF - m' F) = sQ + h(mF - m' F) = sQ.$$

Сравнение выполняется.

Видим, что если $m = m'$ и $m' \leq k_{\max}$, то избиратель проголосовал правильно.

Сложность данного алгоритма формирования и проверки доказательства корректности заполнения бюллетеня в целом можно оценить на основе вышеприведенных соотношений так:

– количество умножений точки эллиптической кривой на число на стороне избирателя – 1М;

– количество умножений точки эллиптической кривой на число в БЧ – 3М.

Рассмотрим примеры формирования и проверки доказательства корректности заполнения всего бюллетеня.

Пример 1. Избиратель правильно заполнил бюллетень.

Пусть, согласно регламенту, избиратель V_i может проголосовать «за» за одного или двух из четырех кандидатов $D1, D2, D3, D4$, и он выбрал двух кандидатов и вычислил криптограммы:

$C_1 = ((19, 9), (34, 48))$, $C_2 = ((49, 10), (34, 11))$,

$C_3 = ((54, 39), (54, 20))$ и $C_4 = ((6, 19), (19, 9))$,

$C_\Sigma = ((51, 57), (49, 49))$, $r = \sum r_i = 10$, $m = \sum m_i = 2$.

Далее он создает доказательство корректности голосования, выполняя следующий алгоритм:

– находит $T = t \cdot Q$, где $t \in Z_p$ случайное число, $t = 2$, $T = t \cdot Q = 2(6, 19) \bmod 59 = (34, 11)$;

– вычисляет хэш-функцию:

$$h = H\left(Q, A_\Sigma, B_\Sigma, T, m\right) = 9;$$

– вычисляет $s = t + r \cdot h; s = 2 + 10 * 9 \bmod 17 = 7$;

– посылает в БЧ $(T = (34, 11), s = 7, m' = 2)$.

Далее БЧ вычисляет $h = H(Q, \sum A_i, \sum B_i, T, m') = 9$, и проверяет сравнение $sQ \stackrel{?}{=} T + h(\sum_{i=1}^{k} B_i - m' F_i)$:

$$sQ \bmod p = 7(6, 19) \bmod 59 = (49, 49);$$

$$T + h\left(\sum_{i=1}^{k} B_i - m' F_i\right) = (34,11) +$$

$$+ 9((49, 49) - 2(54, 39) = (49, 49),$$

$(49, 49) = (49, 49)$: сравнение выполняется.

Пример 2. Избиратель неправильно заполнил бюллетень.

Пусть избиратель V_i проголосовал «за» в пользу трех из четырех кандидатов $D1$, $D2$, $D3$, $D4$, хотя, согласно правилу, он может проголосовать за одного или двух из четырех кандидатов ($k_{\max} = 2$).

Этому выбору соответствуют криптограммы:

$$C_1 = ((19,9),(34,48)),\ C_2 = ((49,10),(6,40)),$$
$$C_3 = ((6,19),(6,19)),\ C_4 = ((54,39),(54,20)),$$
$$C_\Sigma = ((51,57),(6,19)),\ r = \sum r_i = 10,\ \ m = \sum m_i = 3.$$

Избиратель выполняет следующий алгоритм:

– находит $T = t \cdot Q$, где $t \in Z_p$ случайной число, $t = 2, T = t \cdot Q = 2(6,19) \bmod 59 = (34,11)$;

– вычисляет хэш-функцию:

$$h = H\left(Q, A_\Sigma, B_\Sigma, T, m'\right) = 4;$$

– вычисляет $s = t + r \cdot h$; $s = 2 + 10 * 14 \bmod 17 = 6$;

– посылает в БЧ ($T = (34,11)$, $s = 6$, $m' = 2$): избиратель, чтобы скрыть, что он проголосовал неправильно, посылает значение $m' = 2$.

Далее БЧ вычисляет $h = H(Q, \sum A_i, \sum B_i, T, m') = 4$ и проверяет сравнение $sQ \stackrel{?}{=} T + h\left(\sum_{i=1}^{k} B_i - m'F_i\right)$:

$$sQ \bmod p = 6(6,19) \bmod 59 = (11,55);$$

$$T + h\left(\sum_{i=1}^{k} B_i - m'F_i\right) = (34,11) +$$
$$+ 4\left((6,19) - 2(54,39)\right) = (11,4),$$

то есть сравнение $(11,55) \neq (11,4)$ не выполняется. Следовательно, обнаружено некорректное заполнение бюллетеня.

3. Метод проверки корректности заполнения бюллетеня на основе перемешивания криптограмм бюллетеня

Идея этого метода [20] заключается в следующем: сначала сервер генерирует бланк – бюллетень, представляющий вектор C из зашифрованных следующим образом криптограмм:

$$C = (C_1, .., C_k).$$

Первая криптограмма вычисляется как:

$$C_1 = (\rho_1 P, F + \rho_1 Q) \bmod p, \qquad (16)$$

где P – базовая точка; $Q = dP \bmod p$ – открытый ключ, и точка $F = M_i P \bmod p$; $P, Q, F \in E_P(GF(P))$.

Остальные криптограммы вычисляют как:

$$C_i = (\rho_i P, \rho_i Q) \bmod p, \qquad (17)$$

где ρ_i выбирается случайным образом, $\rho_1 \in Z_p$.

Сервер публикует C_i и ρ_i на БЧ.

Избиратель для голосования считывает из БЧ бланк-бюллетень и выполняет следующее:

1) убеждается, что информация, полученная с БЧ, корректна; для этого избиратель проверяет, что $\rho_i P = A_i$ и вычисляет $\mathrm{Rev}_r(C_i) = B_i - \rho_i Q$ – в результате должно получиться либо точка F, либо точка $\mathcal{O}$;

2) приступает к голосованию:

– выбирает своего кандидата – Ds;

– выбирает перестановку $\pi(s, i_1, i_2, \ldots, i_{k-1})$;

– перемешивает C в соответствии с выбранной перестановкой и маскирует бюллетень:

а) генерирует случайным образом набор целых чисел $r_i \in Z_p$;

б) вычисляет:

$$C_i' = C_{\pi(i)} + (r_i P, r_i Q) = \left(A_{\pi(i)} + r_i P, B_{\pi(i)} + r_i Q\right) =$$
$$= \left((\rho_i + r_i)P, F_i + (\rho_i + r_i)Q\right) \bmod p,$$

где $i = 1, 2, \ldots, k$, причем $F_i = \mathcal{O}$ для $i = 2, \ldots, k$.

C_i' отправляет в БЧ;

– формирует доказательство корректности перемешивания бюллетеня, для чего:

– получает от БЧ выбранные случайным образом числа s_i и s_i', $s_i, s_i' \in \{0, 1, .., 2^L - 1\}$;

– вычисляет числа $t_i = s_{\pi(i)}, t_i' = s'_{\pi(i)}$,

– генерирует случайным образом набор целых чисел $r'_i \in Z_p$;

– вычисляет $C_i'' = t_i C_i' + (r_i' P + r_i' Q) = (A_i'', B_i'') = (t_i A_i' + r_i' P, t_i B_i' + r_i' Q)$;

– отправляет C', C'', t_i и t_i' в БЧ.

Проверка доказательства заключается в проверке выполнения сравнений [2, 3, 20]:

$$\sum_{i=1}^{k} \mathrm{Dec}(C_i) \times s_i \stackrel{?}{=} \sum_{i=1}^{k} \mathrm{Dec}(C_i') \times t_i, \qquad (18)$$

$$\sum_{i=1}^{k} \mathrm{Dec}(C_i) \times s_i' \stackrel{?}{=} \sum_{i=1}^{k} \mathrm{Dec}(C_i') \times t_i', \qquad (19)$$

$$\sum_{i=1}^{k} \mathrm{Dec}(C_i) \times s_i \times s_i' \stackrel{?}{=} \sum_{i=1}^{k} \mathrm{Dec}(C_i') \times t_i \times t_i'. \qquad (20)$$

Однако непосредственная проверка согласно (18–20) невозможна, так как для этого БЧ должен знать закрытый ключ d. Поэтому проверка доказательств осуществляется на основе NIZKP. Доказательство (18–20) заключается в проверке следующих равенств [3]:

$$ZP(t_i, r_i' | C_i'' = ((t_i C_i' + (r_i' P, r_i' Q)), \qquad (21)$$

где

$$C_i = (A_i, B_i) = (\rho_i P, F_i + \rho_i Q),$$
$$C_i' = (r_i + r_i')P, F_i + (r_i + r_i')Q,$$
$$C_i'' = t_i A_i' + r_i' P, t_i B_i' + r_i' Q),$$
$$ZP(t_i, r_i' r_i | \sum_{i=1}^{k} (t_i\left(C_i \cdot s_i + (r_i P, r_i Q)\right) + (r_i' P, r_i' Q) = \qquad (22)$$
$$= \sum_{i=1}^{k} C_i'',$$

$$ZP(r_i, r_i', t_i, t_i' | \sum_{i=1}^{k} (t_i'(C_i \cdot s_i' + (r_i P, r_i Q)) = \sum_{i=1}^{k} C_i' t_i', \quad (23)$$

$$ZP(r_i, r_i', t_i, t_i' | \sum_{i=1}^{k} (t_i t_i'(C_i \cdot s_i s_i' + (r_i P, r_i Q)) + \\ + t_i'(r_i' P, r_i' Q) = \sum_{i=1}^{k} C_i'' t_i. \quad (24)$$

Проверку сравнений (21–24) будем проводить отдельно для каждой части криптограммы $C_i'' = (A_i'', B_i'')$,

Для проверки (12) необходимо доказать:

$$A_i'' = A_i' t_i + r_i' P, \ B_i'' = B_i' t_i + r_i' Q.$$

Покажем это для A_i''.

Избиратель формирует доказательство следующем образом:

– выбирает случайные числа $z_i, u_i \in Z_p$, вычисляет:

$$L_i = z_i P \bmod p, \ J_i = u_i A_i' \bmod p \quad (25)$$

и находит хеш-функцию $h = H(A_i'', P, L_i, J_i)$;

– вычисляет:

$$\theta_i = z_i + r_i' h_i \bmod q, \alpha_i = u_i + t_i h_i, \\ T_i = \theta_i P + \alpha_i A_i' \bmod p; \quad (26)$$

– пересылает в БЧ (T_i, L_i, J_i).

БЧ вычисляет хеш-функцию $h' = H(A_i'', P, L_i + J_i)$ и проверяет сравнение: $L_i + J_i + h' \cdot A_i'' =^? T_i$. (27)

Покажем, что если перемешивание выполнено правильно и $h = h'$, то сравнение выполняется. Для этого вычислим левую часть:

$$(L_i + J_i + h \cdot A_i'' = z_i P + u_i \cdot A_i' + h(t_i A_i' + r_i P) = \\ = z_i P + h \cdot r_i' P + u_i \cdot A_i' + t_i \cdot h \cdot A_i' = \theta_i P + \alpha_i A_i'.$$

Видно, что левая часть совпала с правой частью $T_i = \theta_i P + \alpha_i A_i'$. Сравнение (26) для A_i'' доказано.

Затем БЧ проверяет ZP (22) для первых частей криптограмм C_i''.

Избиратель генерирует случайное число $w \in Z_p$. Далее вычисляет:

$$T = wP \bmod p; \quad (28)$$

$$r_\Sigma = \sum_{i=1}^{k} r_i t_i + r_i', U = r_\Sigma P \bmod p; \quad (29)$$

хеш-функцию $h = H(P, T, U, A_1'', A_2'', .., A_k'')$;

$$z = w - r_\Sigma \cdot h \bmod q. \quad (30)$$

После чего отправляет в БЧ (T, z).

БЧ вычисляет:

$$U' = \sum_{i=1}^{k} A_i'' - \sum_{i=1}^{k} s_i A_i; \quad (31)$$

хеш-функцию $h' = H(P, T, U, A_1'', A_2'', .., A_k'')$;

$$T' = zP + h'U'. \quad (32)$$

Если $T = T'$, то (22) для первой части криптограмма C_i'' доказано.

Покажем, что это действительно так:

$$U' = \sum_{i=1}^{k} A_i'' - \sum_{i=1}^{k} s_i A_i = \sum_{i=1}^{k} t_i A_i' + r_i' P - \\ - \sum_{i=1}^{k} s_i A_i = \sum_{i=1}^{k} t_i \left(A_{\pi(i)} + r_i P\right) + r_i' P - \sum_{i=1}^{k} s_i A_i = \\ = \sum_{i=1}^{k} t_i A_{\pi(i)} + \sum_{i=1}^{k} t_i r_i P + r_i' P - \sum_{i=1}^{k} s_i A_i = \\ = \sum_{i=1}^{k} t_i A_{\pi(i)} - \sum_{i=1}^{k} s_i A_i + \sum_{i=1}^{k} (t_i r_i + r_i') P = \\ - \sum_{i=1}^{k} s_{\pi(i)} A_i - \sum_{i=1}^{k} s_i A_i + \left(\sum_{i=1}^{k} t_i r_i + r_i'\right) P.$$

Так как для перестановки π():

$$\sum_{i=1}^{k} s_{\pi(i)} A_{\pi(i)} - \sum_{i=1}^{k} s_i A_i = 0,$$

то

$$U' = \left(\sum_{i=1}^{k} t_i r_i + r_i'\right) P = r_\Sigma P. \quad (33)$$

Далее $T' = zP + h'U' = \left(w - r_\Sigma h\right) P + h' r_\Sigma P$. Если $h' = h$, то $T' = T$.

Аналогично проверяются сравнения (23) и (24).

Заметим, что подсчет голосов в такой системе осуществляется на сервере путем покомпонентного агрегирования координат векторов C_i', полученных от всех избирателей, принявших участие в выборах. В этом случае сумма $\sum_{i=1}^{n} C'_1 = \sum_{i=1}^{n} (C_1 v_i)$ – количество голосов (в зашифрованном виде), поданных за первого кандидата ($v_i = (1, 0)$), $\sum_{i=1}^{n} C'_2 = \sum_{i=1}^{n} (C_1 v_i)$ – количество голосов, поданных за второго кандидата и т. д. Расшифрование агрегированных голосов осуществляется избирательной комиссией с использованием секретного ключа d. На основе гомоморфного свойства схемы шифрования Эль-Гамаля получим расшифровку криптограмм, поданных, например, за i-го кандидата: $\mathrm{Dec}(\sum_{i=1}^{n} C'_i) = R_i$.

Логарифмируя это выражение, найдем сумму голосов (R_1), поданных за i-го кандидата. Победителем на выборах будет кандидат, набравший наибольшую сумму голосов – $\max(R_i)$. В таблицах 4, 5 приведены оценки сложности выполнения проверки корректности заполнения бюллетеня на основе перестановок на стороне избирателя и в БЧ.

Рассмотрим пример формирования и проверки доказательства правильности перемешивания бюллетеня. Пусть избиратель V_i может голосовать только за одного из четырех кандидатов $D1$, $D2$, $D3$, $D4$.

ТАБЛИЦА 4. Оценка сложности метода проверки корректности заполнения бюллетеня на основе перестановок

TABLE 4. Evaluation of the Complexity of the Method of Verifying the Correctness of Filling out the Ballot Based on Permutations

Операции, выполняемые избирателем	Оценка сложности для k-кандидатов
1) Проверка C_i, принятых от БЧ, $r_iP = A_i$ и вычисление $\text{Rev}_r(C_i) = B_i - r_iQ$;	2kM
2) Вычисление: $C_i' = ((r_i + r_i')P, F_i + (r_i + r_i')Q)$ и $C_i'' = (t_iA_i' + r_i'P, t_iB_i' + r_i'Q)$;	6kM
3) Вычисление точек эллиптической кривой $L_i = z_iP, J_i = u_iA_i', T_i = \theta_iP + \alpha_iA_i'$ для доказательства (21) для первой части криптограммы A_i (аналогично для второй части B_i).	8kM
4) Вычисление точек эллиптической кривой $U = r_\Sigma P, T = wP$ для доказательства (21–24).	6M
Всего	16kM + 6M

ТАБЛИЦА 5. Оценка сложности процедуры проверки корректности перемешивания бюллетеня

TABLE 5. Evaluation of the Complexity of the Procedure for Checking the Correctness of Mixing the Ballot

Операции, выполняемые БЧ	Оценка сложности
1) Вычисление левой части сравнения (21) $L_i + J_i + h'A_i'' = T_i$ для доказательства (21)	2kM
2) Вычисление $U' = \sum_{i=1}^{k} A_i'' - \sum_{i=1}^{k} s_iA_i$ и левой части сравнения $zP + h'U' = T$.	1kM 2M
Всего	3kM + 2M

Сервер генерирует бланк – бюллетень, содержащий вектор C из зашифрованных криптограмм: $C = (C_1, C_2, C_3, C_4)$ в соответствии с (16, 17), где F – точка на эллиптической кривой такая же, как в п. 3. $C_i = (\rho_iP, \rho_iQ) \bmod p, i = 2, 3, 4$:

$$C_1 = (\rho_1P, F + \rho_1Q) \bmod p = ((49, 10), (6, 40)),$$
$$C_2 = \big((19, 9), (6, 19)\big), C_3 = \big((54, 39), (54, 20)\big),$$
$$C_4 = \big((6, 19), (19, 9)\big).$$

Сервер публикует:

C = ((49, 10), (6, 40)), ((19, 9), (6, 19)), ((54, 39), (54, 20)), ((6, 19), (19, 9)) и $\rho_i = \{2, 1, 3, 4\}$ в БЧ.

Рассмотрим два случая голосования.

Случай 1. Избиратель правильно перемешал голоса в бюллетене.

Избиратель считывает из БЧ бланк-бюллетень для голосования и убеждается, что информация, полученная от БЧ, корректна. Для этого он выполняет проверки: $\rho_1P = A_i$;$\text{Rev}_r(C_i) = B_i - \rho_1Q = F$.

Для нашего примера

$$C_1 = (A_1, B_1) = ((49, 10), (6, 40)).$$

Проверяет сравнение $A_1 =^? \rho_1P \bmod p$; $(49, 10) = (49, 10)$ и $\text{Rev}_r(C_1) = B_1 - \rho_1Q = (54, 39) = F$.

$$C_2 = (A_2, B_2) = ((19, 9), (6, 19)).$$

Аналогично проверяет $A_2 =^? \rho_2P \bmod p$; $(19, 9) = (19, 9)$; и $\text{Rev}_r(C_2) = B_2 - \rho_2Q = \mathcal{O}$.

$$C_3 = (A_3, B_3) = ((54, 39), (54, 20)).$$

Проверяет $A_3 =^? \rho_3P \bmod p$; $(54, 39) = (54, 39)$ и $\text{Rev}_r(C_3) = B_3 - \rho_3Q = \mathcal{O}$.

$$C_4 = (A_4, B_4) = ((6, 19), (19, 9)).$$

Проверяет $A_4 =^? \rho_4P \bmod p$; $(6, 19) = (6, 19)$ и $\text{Rev}_r(C_4) = B_4 - \rho_4Q = \mathcal{O}$.

Все проверки выполнены правильно.

Далее избиратель приступает к голосованию. Во-первых, выбирает своего кандидата – $D4$. Во-вторых, выбирает перестановку: $\pi(1) = 4, \pi(2) = 2, \pi(3) = 3, \pi(4) = 1$. Для этого перемешивает координаты C в соответствии с выбранной перестановкой (таблица 6) и осуществляет маскировку бюллетеня:

– генерирует случайным образом набор целых чисел $r_i = \{4, 2, 1, 3\}$;

– вычисляет $C_i' = C_{\pi(i)} + (r_iP, r_iQ) = (A_{\pi(i)} + r_iP, B_{\pi(i)} + r_iQ) = ((\rho_1 + r_i)P, F_i + (\rho_1 + r_i)Q) \bmod p$, получает:

$\{C_i'\}$ = {((11, 4), (11, 55)), ((54, 39), (54, 20)), ((6, 19), (19, 9)), ((34, 48), (19, 9))}.

В-третьих, формирует доказательство корректности перемешивания бюллетеня. Для этого избиратель получает от БЧ случайным образом выбранные им числа s_i и s_i', где $i = 1, .., 4$. Пусть ($s_1 = 2, s_2 = 3, s_3 = 1, s_4 = 4, s_1' = 1, s_2' = 3, s_3' = 2, s_4' = 4$. Далее избиратель вычисляет числа $t_i = s_{\pi(i)}$, $t_i' = s_{\pi(i)}', i = 1, 2, \dots, k$. Тогда $t_1 = 4, t_2 = 1, t_3 = 3, t_4 = 2$. $t_1' = 4, t_2' = 2, , t_3' = 3, t_4' = 1$. А потом выбирает $r_i' = \{1, 3, 4, 2\}$ и вычисляет:

$$C_i'' = (t_iA_i' + r_i'P, t_iB_i' + r_i'Q),$$

$\{C_i''\}$ = {((6, 40), (19, 50)), ((11, 55), (11, 4)), ((19, 9), (6, 19)), ((19, 9), (51, 57))}

и отправляет C_i', C_i'', t_i и t_i' в БЧ.

ТАБЛИЦА 6. Избиратель правильно перемешал свой бюллетень

TABLE 6. The Voter Shuffled His Ballot Correctly

Кандидаты	$D1$	$D2$	$D3$	$D4$
Криптограммы, составляющие C_i	$C4$	$C2$	$C3$	$\boldsymbol{C1}$

Далее проверим выполнение сравнения (21). Для первой части криптограммы C_1', необходимо доказать, что $A_1'' = A_1't_1 + r_1'P$. В нашем примере мы получили:

$$C_1' = (r_1 + r_1')P, F + (r_1 + r_1')Q = ((11, 4), (11, 5));$$
$$C_1'' = (t_1A_1' + r_1'P, t_1B_1' + r_1'Q) = ((6, 40), (19, 50)).$$

Избиратель формирует доказательство для этой криптограммы следующим образом:

1) выбирает случайные числа $z_1 = 2, u_1 = 1$, вычисляет $L_1 = z_1P \bmod p = (49{,}10), J_1 = u_1A_1' \bmod p = (54{,}20)$ и находит хеш-функцию:

$h_1 = H\big((6, 40), (19, 9), (19, 50)\big) \bmod 17 = 16$;

2) вычисляет:
$\theta_1 = z_1 + r_1' h_1 \bmod q = 2 + 1 * 16 \bmod 17 = 1;$
$\alpha_1 = u_1 + t_1 h_1 \bmod q = 1 + 4 * 16 \bmod 17 = 14;$
$T_1 = \theta_1 P + \alpha_1 A_1' \bmod p = 1(19,9) + 14(11,4) \bmod 59 = (6,40);$

3) пересылает в БЧ ($T_1 = (6,40)$, $L_1 = (49,10)$, $J_1 = (54,20)$).

БЧ вычисляет хеш-функцию $h' = H(A_1'' = (6,40), P = (19,9), L_1 + J_1 = (19,50)) \bmod 17 = 16$ и h_i', а также проверяет сравнение $L_1 + J_1 + h' \cdot A_1'' =? T_1$; $(6,40) = T_1$. Таким образом доказано, что сравнение (21) выполняется для первой части криптограммы C_1.

Аналогично проверяем доказательство (21) для A_2, A_3, A_4.

Проверим выполнение *ZP* (22). Избиратель выполняет следующие действия:

– генерирует случайное число $w = 3$;

– вычисляет $T = wP \bmod p = (54,39)$;

– вычисляет
$r_\Sigma = \sum_{i=1}^{k} r_i t_i + r_i' = 3$, $U = r_\Sigma P \bmod p = (54,39)$;

– вычисляет хеш-функцию $h = H(P = (19,9), T = (54,39), U = (54,39), A_1'' = (6,40), A_2'' = (11,55), A_3'' = (19,9), A_4'' = (19,9)) \bmod 17 = 10$;

– вычисляет $z = w - r_\Sigma \cdot h \bmod q = 6$;

– посылает в БЧ ($T = (54,39), z = 6$).

Далее БЧ вычисляет:

$$U' = \sum_{i=1}^{k} A_i'' - \sum_{i=1}^{k} s_i A_i = \left(\sum_{i=1}^{k} t_i r_i + r_i'\right) P = r_\Sigma P = (54,39);$$

– хеш-функцию $h' = 10$;

– $T' = zP + h'U' = (54,39)$;

– проверяет $T =^? T'$; $(54,39) = (54,39)$, т. е. (22) для первой части криптограмм C_i'' доказано.

Случай 2. Избиратель неправильно перемешал голоса в бюллетене.

Все шаги выполняются, как в предыдущем примере, до момента перемешивания. Избиратель выполняет перестановку π, как показано в таблице 7.

ТАБЛИЦА 7. Избиратель неправильно перемешал свой бюллетень

TABLE 7. The Voter Mixed His Ballot Incorrectly

Кандидаты	*D1*	*D2*	*D3*	*D4*
Криптограммы, составляющие C_1	*C1*	***C1***	*C3*	*C4*

$\{C_i\} = \{((19,9),(6,19)), ((19,9),(6,19)), ((54,39),(54,20)), ((6,19),(19,9))\}$.

Далее осуществляет маскировку бюллетеня, генерирует случайным образом набор целых чисел $r_i = \{4, 2, 1, 3\}$, вычисляет C_i':

$$C_i' = C_{\pi(i)} + (r_i P, r_i Q) = (A_{\pi(i)} + r_i P, B_{\pi(i)} + r_i Q) = ((\rho_i + r_i)P, F_i + (\rho_i + r_i)Q) \bmod p,$$

для $i = 2, .., k$, получает:

$\{C_i'\} = \{((11,4),(54,20)), ((54,39),(11,4)), ((6,19),(19,9)), ((34,48),(19,9))\}$.

Следующий шаг – формирование доказательства корректности перемешивания бюллетеня. Для этого избиратель получает от БЧ случайным образом выбранные числа s_i и s_i', где $i = 1, .., n$: пусть $(s_1 = 2, s_2 = 3, s_3 = 1, s_4 = 4, s_1' = 1, s_2' = 3, s_3' = s_3' = 2, s_4' = 4)$, вычисляет числа t_i, t_i', как в предыдущем примере, выбирает $r_i' = \{1, 2, 3, 4\}$ и вычисляет:

$$C_i'' = (t_i A_i' + r_i' P, t_i B_i' + r_i' Q),$$

$\{C_i''\} = \{((6,40),(19,50)), ((11,55),(11,4)), ((19,9),(6,19)), ((19,9),(51,57))\}$,

после чего отправляет $C_i', C_i'' =$, t_i и t_i' в БЧ.

Далее проверим доказательства (21) для первых частей криптограммы C_1', для этого покажем, что $A_1'' = A_1' t_1 + r_1' P$.

Ранее было получено:

$C_1' = (r_1 + r_1')P$, $F + (r_1 + r_1')Q = ((11,4),(54,20))$ и $C_1'' = (t_1 A_1' + r_1' P, t_1 B_1' + r_1' Q) = ((6,40),(19,50))$.

Избиратель формирует доказательство для каждой криптограммы:

– выбирает случайные числа $z_1 = 2, u_1 = 1$, вычисляет $L_1 = z_1 P \bmod p = (49,10)$, $J_1 = u_1 A_1' \bmod p = (54,20)$ и хеш-функцию $h_1 = H(A_1'' = (6,40), P = (19,9), L_1 + J_1 = (19,50)) \bmod 17 = 16$;

– вычисляет:
$\theta_1 = z_1 + r_1' h_1 \bmod q = 2 + 1 * 16 \bmod 17 = 1;$
$\alpha_1 = u_1 + t_1 h_1 \bmod q = 1 + 4 * 16 \bmod 17 = 14;$
$T_1 = \theta_1 P + \alpha_1 A_1' \bmod p = 1(19,9) + 14(11,4) \bmod 59 = (6,40);$

– пересылает в БЧ ($T_1 = (6,40), P = (19,9), L_1 + J_1 = (19,50)$).

БЧ вычисляет хеш-функцию:

$$h' = H(A_1' = (6,40), P = (19,9), L_1 + J_1 = (19,50)) \bmod 17 = 16$$

и проверяет сравнение $L_1 + J_1 + h' \cdot A_1'' =^? T_1$; $(6,40) = T_1$. Сравнение выполняется, т. е. (21) для первой части криптограммы C_1 доказано.

Аналогично проверяем доказательство (21) для A_2, A_3, A_4.

Проверим доказательство *ZP* (22).

Избиратель генерирует случайное число $w = 3$. После чего вычисляет:

– $T = wP \bmod p = (54,39)$;

– $r_\Sigma = \sum_{i=1}^{k} r_i t_i + r_i' = 3, U = r_\Sigma P \bmod p = (54,39)$;

– хеш-функцию $h = H(P = (19,9), T = (54,39), U = (54,39), A_1'' = (6,40), A_2'' = (11,55), A_3'' = (19,9), A_4'' = (19,9)) \bmod 17 = 10$;

– $z = w - r_\Sigma \cdot h \bmod q = 6$.

Далее посылает в БЧ ($T = (54,39), z = 6$).

БЧ вычисляет:
– $U' = \sum_{i=1}^{k} A_i'' - \sum_{i=1}^{k} s_i A_i = (51, 57)$,
– хеш-функцию $h' = 10$;
– $T' = zP + h'U' = (6, 40)$.
– проверяет $T =^{?} T'$; $(54, 39) \neq (6, 40)$ – не выполнено; следовательно, сравнение (22) для первой части криптограммы C_1'' не доказано.

4. Сравнительный анализ сложности реализации методов доказательства корректности заполнения бюллетеня избирателем

Проведем анализ сложности реализации рассмотренных выше методов проверки корректности заполнения бюллетеня избирателем в системе ДЭГ, основанных на сравнении дискретных логарифмов и на проверке корректности перестановки. Будем полагать, что в обоих случаях для шифрования используются криптосистемы Эль-Гамаля на эллиптической кривой с одинаковыми параметрами (уравнение кривой, длины ключей, длины криптограмм, длины случайных чисел). Результаты сравнения представлены в таблице 8. Сложность вычислений будем оценивать количеством выполнений наиболее сложной операции умножения точки на целое число (буква М). Оценку сложности проведем отдельно для избирателя (доказывающей стороны) и БЧ (проверявшей стороны).

Таблица 8 показывает, что сложность формирования доказательства корректности заполнения бюллетеня для k кандидатов составляет $10k$М + 1М операций умножения для первого метода и $16k$М + 6М – для второго; это примерно на 60 % меньше для первого метода. Наоборот, объем вычислений для проверки доказательства корректности заполнения бюллетеня на одного избирателя, проводимых в БЧ, составляет $5k$М + 3М операций умножения для первого метода и $3k$М + 2М для второго. Т. е. на проверку бюллетеня во втором методе требуется в 1,67 раза меньше вычислений, чем в первом. Этот выигрыш существенно возрастает с увеличением количества избирателей, что показано в таблице 9.

Можно сделать вывод, что при большом количестве избирателей предпочтительным методом проверки является перемешивание зашифрованных голосов у избирателя, так как в этом случае значительно уменьшается нагрузка на БЧ, связанная с проверкой корректности заполнения бюллетеней.

ТАБЛИЦА 8. Сравнение методов доказательства корректности заполнения бюллетеня избирателем

TABLE 8. Comparison of Methods for Confirming the Correctness of Filling out the Voter's Ballot

	Метод на основе	
	сравнения дискретных логарифмов	проверки корректности перестановки
1) Количество операций, выполняемых на стороне избирателями. Шифрование бюллетеня	$3k$М	–
2) Количество операций формирования доказательства избирателем	$7k$М + 5М	$16k$М + 6М
Всего на стороне избирателя	$10k$М + 1М	$16k$М + 6М
Формирование зашифрованных криптограмм		$4k$М (один раз для всех избирателей)
3) Общее количество операций для проверки доказательства в БЧ	$5k$М + 3М	$3k$М + 2М
Всего на стороне БЧ для одного избирателя	$5k$М + 3М	$3k$М + 2М
Всего на стороне БЧ для n избирателей	n ($5k$М + 3М)	n ($3k$М + 2М) + $4k$М

ТАБЛИЦА 9. Оценка сложности вычислений в БЧ для первого и второго методов для разного количества избирателей

TABLE 9. Evaluation the Complexity of Calculations in the BC for the First and Second Methods for Different Numbers of Voters

	n = 1	n = 10	n = 100	n = 1000	n = 10000	n = 100000
k = 3						
Метод 1	28	280	2800	28000	2850000	2800000
Метод 2	23	122	1112	11012	110012	1100012
k = 4						
Метод 1	18	180	1800	18000	180000	1800000
Метод 2	30	156	1416	14016	140016	1400016
k = 5						
Метод 1	23	230	2300	235000	230000	2300000
Метод 2	37	190	1720	17020	170020	1700020
k = 10						
Метод 1	53	530	5300	53000	530000	5300000
Метод 2	72	360	3240	32040	320040	3200040

Заключение

В работе рассмотрена система ДЭГ, построенная на основе гомоморфной криптосистемы Эль-Гамаля на эллиптической кривой. Рассмотрены методы защиты системы ДЭГ от угрозы со стороны избирателя, заключающейся в неправильном заполнении бюллетеня. Нетривиальность решения этой задачи состоит в том, что нужно определить корректность заполнения бюллетеня избирателем, представленного в зашифрованном виде, т. е. без ознакомления с решением, которое сделал избиратель, выбирая кандидатов.

Исследованы два метода проверки корректности заполнения бюллетеня, основанные на применении доказательств с нулевым разглашением секрета. Приведено детальное описание обоих методов, подкрепленное примерами правильного и неправильного заполнения бюллетеня. Оценена сложность реализации методов по количеству операций умножения точки эллиптической кривой на целое число. Сравнительный анализ показал, что первый метод требует меньшего объема вычислений у избирателя. Для второго метода, наоборот, количество операций умножения, проводимых при проверке доказательства, примерно в 1,67 раза меньше, чем для первого.

Можно дать такие рекомендации по применению этих методов. Во-первых, следует принять во внимание, что сложность доказательства корректности заполнения бюллетеня требует во много раз большего количества операций по сравнению с операциями шифрования и расшифрования голоса избирателя, которые принципиально необходимы для обеспечения тайны голосования. Во-вторых, при выборе метода необходимо учитывать масштабность системы ДЭГ. При малом количестве избирателей исследуемые методы примерно равноценны по сложности вычислений. При большом количестве избирателей (более 10000) второй метод предпочтительней.

Также следует учесть, что второй метод имеет преимущество в том, что избиратель сам не шифрует свой бюллетень, а устанавливает зашифрованную метку в бланке бюллетеня на позицию выбираемого кандидата, затем маскирует голоса и перемешивает бюллетень.

Список источников

1. Furukawa J., Mori K., Sako K. An implementation of a Mix-Net Based Network Voting Scheme and Its Use in a Private Organization // In Chaum D., Jakobsson M., Rivest R.L., Ryan P.Y.A., Benaloh J., Kutylowski M., Adida B. ed. Towards Trustworthy Elections. Lecture Notes in Computer Science. Vol. 6000. Berlin, Heidelberg: Springer, 2010. PP. 141–154. DOI:10.1007/978-3-642-12980-3_8

2. Peng K. An efficient shuffling based eVoting scheme // Journal of Systems and Software. 2011. Vol. 84. № 6. PP. 906–922. DOI:10.1016/j.jss.2011.01.001

3. Peng K., Dawson E., Bao F. Modification and optimisation of a shuffling scheme: Stronger security, formal analysis and higher efficiency // International Journal of Information Security. 2011. Vol. 10. PP. 33–47. DOI:10.1007/s10207-010-0117-y

4. Adida B. Helios: Web-based Open-Audit Voting // Proceedings of the 17th USENIX Security Symposium (San Jose, USA, 28 July–1 August 2008). 2008. PP. 335–348.

5. Fujioka A., Okamoto T., Ohta K. A practical secret voting scheme for large scale elections // Proceedings of the Workshop on the Theory and Application of Cryptographic Techniques «Advances in Cryptology – AUSCRYPT '92» (Gold Coast, Australia, 13–16 December 1992). Lecture Notes in Computer Science. Vol. 718. Berlin, Heidelberg: Springer, 1993. PP. 245–251. DOI:10.1007/3-540-57220-1_66

6. Ibrahim S., Kamat M., Salleh M., Aziz S.R.A. Secure E-voting with blind signature // Proceedings of the 4th National Conference of Telecommunication Technology, NCTT 2003, Shah Alam, Malaysia, 14–15 January 2003). IEEE, 2003. PP. 193–197. DOI:10.1109/NCTT.2003.1188334

7. Mateu V., Sebé F., Valls M. Constructing credential-based E-voting systems from offline E-coin protocols // Journal of Network and Computer Applications. 2014. Vol. 42. PP. 39–44. DOI:10.1016/j.jnca.2014.03.009

8. Killer C., Rodrigues B., Scheid E.J., Franco M., Eck M., Zaugg N., et al. Provotum: A Blockchain-based and End-to-end Verifiable Remote Electronic Voting System // Proceedings of the 45th Conference on Local Computer Networks (LCN, Sydney, Australia, 16–19 November 2020). IEEE, 2020. PP. 172–183. DOI:10.1109/LCN48667.2020.9314815

9. Aziz A.A., Qunoo H.N., Samra A.A. Using Homomorphic Cryptographic Solutions on E-voting Systems // International Journal of Computer Network and Information Security. 2018. Vol. 12. Iss. 1. PP. 44–59. DOI:10.5815/ijcnis.2018.01.06

10. Yang X., Yi X., Nepal S., Kelarev A., Han F. A secure Verifiable Ranked Choice Online Voting System Based on Homomorphic Encryption // IEEE Access. 2018. Vol. 6. PP. 20506–20519. DOI:10.1109/ACCESS.2018.2817518

11. ElGamal T. A public key cryptosystem and a signature scheme based on discrete logarithms. Springer – Verlag, 1998. URL: https://people.csail.mit.edu/alinush/6.857-spring-2015/papers/elgamal.pdf (дата обращения 10.04.2023)

12. Alonso L.P., GASCÓ M., del BLANCO D.Y.M., Alonso J.Á.H., Barrat J., Moreton H.A. E-Voting System Evaluation Based on the Council of Europe Recommendations: Helios Voting // IEEE Transactions on Emerging Topics in Computing. 2021. Vol. 9. Iss. 1. PP. 161–173. DOI:10.1109/TETC.2018.2881891

13. Balzarotti D., Banks G., Cova M., Felmetsger V., Kemmerer R., Robertson W., et al. An experience in Testing the Security of Real-World Electronic Voting Systems // IEEE Transactions on Software Engineering. 2010. Vol. 36. Iss. 4. PP. 453–473. DOI:10.1109/TSE.2009.53

14. Estehghari S., Desmedt Y. Exploiting the Client Vulnerabilities in Internet E-voting Systems: Hacking Helios 2.0 as an Example // Proceedings of the Electronic Voting Technology Workshop / Workshop on Trustworthy Elections (Washington,

USA, 9–10 August 2010). 2010.

15. Butterfield K., Zou X. Analysis and Implementation of Internet Based Remote Voting // Proceedings of the 11th International Conference on Mobile Ad Hoc and Sensor Systems (Philadelphia, USA, 28–30 October 2014). IEEE, 2014. DOI:10.1109/MASS.2014.134

16. Cramer R., Franklin M., Schoenmakers B., Yung M. Multi-Authority Secret-Ballot Elections with Linear Work // Proceedings of the International Conference on the Theory and Application of Cryptographic Techniques (Saragossa, Spain, 12–16 May 1996). «Advances in Cryptology – EUROCRYPT '96». Lecture Notes in Computer Science. Vol. 1070. Berlin, Heidelberg: Springer, 1996. PP. 72–83. DOI:10.1007/3-540-68339-9_7

17. Seol S., Kim H., Park J.H. An Efficient Open Vote Network for Multiple Candidates // IEEE Access. 2022. Vol. 10. PP. 124291–124304. DOI:10.1109/ACCESS.2022.3224798

18. Hao F., Ryan P.Y.A., Zieliński P. Anonymous voting by two-round public discussion // IET Information Security. 2010. Vol. 4. Iss. 2. PP. 62–67. DOI:10.1049/iet-ifs.2008.0127

19. Cramer R., Gennaro R., Schoenmakers B. A Secure and Optimally Efficient Multi-Authority Election Scheme // Proceedings of the International Conference on the Theory and Application of Cryptographic Techniques (Konstanz, Germany, 11–15 May 1997) «Advances in Cryptology – EUROCRYPT '97». Lecture Notes in Computer Science. Vol. 1233. Berlin, Heidelberg: Springer, 1997. PP. 103–118. DOI:10.1007/3-540-69053-0_9

20. Mateu V., Miret J.M., Sebé F. A hybrid approach to vector-based homomorphic tallying remote voting // International Journal of Information Security. 2016. Vol. 15. Iss. 2. PP. 211–221. DOI:10.1007/s10207-015-0279-8

21. Peng K. A general and efficient countermeasure to relation attacks in mix-based e-voting // International Journal of Information Security. 2011. Vol. 10. Iss. 1. PP. 49–60. DOI:10.1007/s10207-010-0122-1

22. Mohr A. A Survey of Zero-Knowledge Proofs with Applications to Cryptography. 2007. URL: http://austinmohr.com/Work_files/zkp.pdf (дата обращения 10.04.2023)

23. Blum M., Feldman P., Micali S. Non-interactive zero-knowledge and its applications // Proceedings of the 12-th annual ACM symposium on Theory of computing (Chicago, USA, 2–4 May 1988). ACM, 1988. PP. 103–112. DOI:10.1145/62212.62222

24. Huqing W., Zhixin S. Research on Zero-Knowledge Proof Protocol // International Journal of Computer Science Issues. 2013. Vol. 10. Iss. 1. PP. 194–200.

25. Feldman P. A practical scheme for non-interactive verifiable secret sharing // Proceedings of the 28th Annual Symposium on Foundations of Computer Science (Los Angeles, USA, 12–14 October 1987). IEEE, 1987. PP. 427–437. DOI:10.1109/SFCS.1987.4

26. Blum M., De Santis A., Micali S., Persiano G. Noninteractive Zero-Knowledge // SIAM Journal on Computing. 1991. Vol. 20. Iss. 6. DOI: 10.1137/0220068

27. Boruah D., Saikia M. Implementation of ElGamal Elliptic Curve Cryptography over prime field using C // Proceedings of the International Conference on Information Communication and Embedded Systems (ICICES2014, Chennai, India, 27–28 February 2014). IEEE, 2014. DOI:10.1109/ICICES.2014.7033751

28. Kapoor V., Abraham V.S., Singh R. Elliptic curve cryptography // Ubiquity. 2008. Vol. 9. Iss. 20. DOI:10.1145/1378355.1378356

29. Коржик В.И., Яковлев В.А. Основы криптографии. СПб.: ИЦ Интермедия, 2016. 296 с.

References

1. Furukawa J., Mori K., Sako K. An implementation of a Mix-Net Based Network Voting Scheme and Its Use in a Private Organization. *In Chaum D., Jakobsson M., Rivest R.L., Ryan P.Y.A., Benaloh J., Kutylowski M., Adida B. ed. Towards Trustworthy Elections. Lecture Notes in Computer Science. Vol. 6000*. Berlin, Heidelberg: Springer; 2010. p.141–154. DOI:10.1007/978-3-642-12980-3_8

2. Peng K. An efficient shuffling based eVoting scheme. *Journal of Systems and Software*. 2011;84(6):906–922. DOI:10.1016/j.jss.2011.01.001

3. Peng K., Dawson E., Bao F. Modification and optimisation of a shuffling scheme: Stronger security, formal analysis and higher efficiency. *International Journal of Information Security*. 2011;10:33–47. DOI:10.1007/s10207-010-0117-y

4. Adida B. Helios: Web-based Open-Audit Voting. *Proceedings of the 17th USENIX Security Symposium, 28 July–1 August 2008, San Jose, USA*. 2008. p.335–348.

5. Fujioka A., Okamoto T., Ohta K. A practical secret voting scheme for large scale elections. *Proceedings of the Workshop on the Theory and Application of Cryptographic Techniques «Advances in Cryptology – AUSCRYPT '92», 13–16 December 1992, Gold Coast, Australia. Lecture Notes in Computer Science, vol.718*. Berlin, Heidelberg: Springer; 1993. p.245–251. DOI:10.1007/3-540-57220-1_66

6. Ibrahim S., Kamat M., Salleh M., Aziz S.R.A. Secure E-voting with blind signature. *Proceedings of the 4th National Conference of Telecommunication Technology, NCTT 2003, 14–15 January 2003, Shah Alam, Malaysia*. IEEE; 2003. p.193–197. DOI:10.1109/NCTT.2003.1188334

7. Mateu V., Sebé F., Valls M. Constructing credential-based E-voting systems from offline E-coin protocols. *Journal of Network and Computer Applications*. 2014;42:39–44. DOI:10.1016/j.jnca.2014.03.009

8. Killer C., Rodrigues B., Scheid E.J., Franco M., Eck M., Zaugg N., et al. Provotum: A Blockchain-based and End-to-end Verifiable Remote Electronic Voting System. *Proceedings of the 45th Conference on Local Computer Networks, LCN, 16–19 November 2020, Sydney, Australia*. IEEE; 2020. p.172–183. DOI:10.1109/LCN48667.2020.9314815

9. Aziz A.A., Qunoo H.N., Samra A.A. Using Homomorphic Cryptographic Solutions on E-voting Systems. *International Journal of Computer Network and Information Security*. 2018;12(1):44–59. DOI:10.5815/ijcnis.2018.01.06

10. Yang X., Yi X., Nepal S., Kelarev A., Han F. A secure Verifiable Ranked Choice Online Voting System Based on Homomorphic Encryption. *IEEE Access*. 2018;6:20506–20519. DOI:10.1109/ACCESS.2018.2817518

11. ElGamal T. *A public key cryptosystem and a signature scheme based on discrete logarithms*. Springer – Verlag; 1998. URL: https://people.csail.mit.edu/alinush/6.857-spring-2015/papers/elgamal.pdf [Accessed 10th April 2023]
12. Alonso L.P., GASCÓ M., del BLANCO D.Y.M., Alonso J.Á.H., Barrat J., Moreton H.A. E-Voting System Evaluation Based on the Council of Europe Recommendations: Helios Voting. *IEEE Transactions on Emerging Topics in Computing*. 2021;9(1):161–173. DOI:10.1109/TETC.2018.2881891
13. Balzarotti D., Banks G., Cova M., Felmetsger V., Kemmerer R., Robertson W., et al. An experience in Testing the Security of Real-World Electronic Voting Systems. *IEEE Transactions on Software Engineering*. 2010;36(4):453–473. DOI:10.1109/TSE.2009.53
14. Estehghari S., Desmedt Y. Exploiting the Client Vulnerabilities in Internet E-voting Systems: Hacking Helios 2.0 as an Example. *Proceedings of the Electronic Voting Technology Workshop / Workshop on Trustworthy Elections, 9–10 August 2010, Washington, USA*. 2010.
15. Butterfield K., Zou X. Analysis and Implementation of Internet Based Remote Voting. *Proceedings of the 11th International Conference on Mobile Ad Hoc and Sensor Systems, 28–30 October 2014, Philadelphia, USA*. IEEE; 2014. DOI:10.1109/MASS.2014.134
16. Cramer R., Franklin M., Schoenmakers B., Yung M. Multi-Authority Secret-Ballot Elections with Linear Work. *Proceedings of the International Conference on the Theory and Application of Cryptographic Techniques, 12–16 May 1996, Saragossa, Spain. «Advances in Cryptology – EUROCRYPT '96». Lecture Notes in Computer Science, vol.1070*. Berlin, Heidelberg: Springer; 1996. p.72–83. DOI:10.1007/3-540-68339-9_7
17. Seol S., Kim H., Park J.H. An Efficient Open Vote Network for Multiple Candidates. *IEEE Access*. 2022;10:124291–124304. DOI:10.1109/ACCESS.2022.3224798
18. Hao F., Ryan P.Y.A., Zieliński P. Anonymous voting by two-round public discussion. *IET Information Security*. 2010;4(2):62–67. DOI:10.1049/iet-ifs.2008.0127
19. Cramer R., Gennaro R., Schoenmakers B. A Secure and Optimally Efficient Multi-Authority Election Scheme. *Proceedings of the International Conference on the Theory and Application of Cryptographic Techniques, 11–15 May 1997, Konstanz, Germany. «Advances in Cryptology – EUROCRYPT '97». Lecture Notes in Computer Science, vol.1233*. Berlin, Heidelberg: Springer, 1997. p.103–118. DOI:10.1007/3-540-69053-0_9
20. Mateu V., Miret J.M., Sebé F. A hybrid approach to vector-based homomorphic tallying remote voting. *International Journal of Information Security*. 2016;15(20):211–221. DOI:10.1007/s10207-015-0279-8
21. Peng K. A general and efficient countermeasure to relation attacks in mix-based e-voting. *International Journal of Information Security*. 2011;10(1):49–60. DOI:10.1007/s10207-010-0122-1
22. Mohr A. *A Survey of Zero-Knowledge Proofs with Applications to Cryptography*. 2007. URL: http://austinmohr.com/Work_files/zkp.pdf [Accessed 10th April 2023]
23. Blum M., Feldman P., Micali S. Non-interactive zero-knowledge and its applications. *Proceedings of the 12-th annual ACM symposium on Theory of computing, 2–4 May 1988, Chicago, USA*. ACM; 1988. p.103–112. DOI:10.1145/62212.62222
24. Huqing W., Zhixin S. Research on Zero-Knowledge Proof Protocol. *International Journal of Computer Science Issues*. 2013;10(1):194–200.
25. Feldman P. A practical scheme for non-interactive verifiable secret sharing. *Proceedings of the 28th Annual Symposium on Foundations of Computer Science, 12–14 October 1987, Los Angeles, USA*. IEEE; 1987. p.427–437. DOI:10.1109/SFCS.1987.4
26. Blum M., De Santis A., Micali S., Persiano G. Noninteractive Zero-Knowledge. *SIAM Journal on Computing*. 1991;20(6). DOI: 10.1137/0220068
27. Boruah D., Saikia M. Implementation of ElGamal Elliptic Curve Cryptography over prime field using C. *Proceedings of the International Conference on Information Communication and Embedded Systems, ICICES2014, 27–28 February 2014, Chennai, India)*. IEEE; 2014. DOI:10.1109/ICICES.2014.7033751
28. Kapoor V., Abraham V.S., Singh R. Elliptic curve cryptography. *Ubiquity*. 2008;9(20). DOI:10.1145/1378355.378356
29. Korzhik V.I., Yakovlev V.A. *Fundamentals of Cryptography*. Saint Petersburg: IC Intermedia Publ.; 2016. 296 p. (in Russ.)

Статья поступила в редакцию 16.03.2023; одобрена после рецензирования 21.03.2023; принята к публикации 24.03.2023.
The article was submitted 16.03.2023; approved after reviewing 21.03.2023; accepted for publication 24.03.2023.

Информация об авторах:

ЯКОВЛЕВ Виктор Алексеевич — доктор технических наук, профессор, профессор кафедры защищенных систем связи Санкт-Петербургский государственный университет телекоммуникаций им. проф. М.А. Бонч-Бруевича
https://orcid.org/0009-0007-2861-9605

САЛМАН Васан Давуд — аспирант кафедры защищенных систем связи Санкт-Петербургский государственный университет телекоммуникаций им. проф. М.А. Бонч-Бруевича
https://orcid.org/0000-0003-4454-7844

Выходные данные

Товарный знак №929373, правообладатель: Федеральное государственное бюджетное образовательное учреждение высшего образования «Санкт-Петербургский государственный университет телекоммуникаций им. проф. М.А. Бонч-Бруевича»,
191186, Санкт-Петербург, наб. реки Мойки, 61, литера А (RU)
Зарегистрирован в Государственном реестре товарных знаков и знаков обслуживания Российской Федерации 13.03.2023 г. Заявка №2022733914

План издания научной литературы 2023 г., п. 7

Дата выхода в свет	Усл.-печ. л.	Формат	Тираж	Заказ	Свободная цена
31.05.2023	13	$60\times84_{1/8}$	1000 экз.	№ 1401	

Ответственный редактор **Татарникова И.М.**
Выпускающий редактор **Яшугин Д.Н.**
Дизайн: **Коровин В.М.**

Адрес типографии:
196105, Санкт-Петербург, Московский пр., д. 149

Учредитель и издатель:
Федеральное государственное бюджетное образовательное учреждение высшего образования "Санкт-Петербургский государственный университет телекоммуникаций им. проф. М.А. Бонч-Бруевича"
E-mail: tuzs@spbgut.ru Web: tuzs.sut.ru VK: vk.com/spbtuzs

Подписной индекс в Объединенном каталоге "ПРЕССА РОССИИ" - 59983