№5-2024

ISSN 1999-9429

# ИЗВЕСТИЯ ЮФУ

## ТЕХНИЧЕСКИЕ НАУКИ

- Алгоритмы обработки информации
- Анализ данных и моделирование
- Электроника, приборостроение и радиотехника

# ИЗВЕСТИЯ ЮФУ. ТЕХНИЧЕСКИЕ НАУКИ
# IZVESTIYA SFedU. ENGINEERING SCIENCES

*Свидетельство о регистрации средства массовой информации*
*ПИ № ФС77-28889 от 12.07.2007*
*Федеральная служба по надзору в сфере массовых коммуникаций, связи*
*и охраны культурного наследия*
*Научно-технический и прикладной журнал*
*Издается с 1995 года, до середины 2007 года под названием «Известия ТРТУ»*
*Подписной индекс ПС704*

*№ 5 (241). 2024 г.*

---

**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».**



***Учредитель*** Южный федеральный университет.
***Издатель*** Южный федеральный университет.
***Ответственный за выпуск*** Самойлов А.Н.
***Технический редактор*** Ярошевич Н.В.
***Оригинал-макет выполнен*** Ярошевич Н.В.

Дата выхода в свет 13.11. 2024 г. Формат $70\times108\ {}^{1}\!/_{16}$. Бумага офсетная.

Офсетная печать. Усл. печ. л. – 24,9. Уч.-изд. л. – 17,5.
Заказ № 9711. Тираж 250 экз.

***Адрес издателя:*** 344090, г. Ростов-на-Дону, пр. Стачки, 200/1, тел. 8(863)243-41-66.

***Адрес типографии:*** Отпечатано в отделе полиграфической, корпоративной и сувенирной продукции Издательско-полиграфического комплекса КИБИ МЕДИА ЦЕНТРА ЮФУ. 344090, г. Ростов-на-Дону, пр. Стачки, 200/1, тел. 8(863)243-41-66.

***Адрес редакции:*** 347922, г. Таганрог, ул. Чехова, 22, ЮФУ, тел. +7 (928) 909-57-82, e-mail: iborodyanskiy@sfedu.ru, http://izv-tn.tti.sfedu.ru/.

16+ Цена свободная

ISSN 1999-9429 (Print)
ISSN 2311-3103 (Online)

## СОДЕРЖАНИЕ

### РАЗДЕЛ I. АЛГОРИТМЫ ОБРАБОТКИ ИНФОРМАЦИИ

**Л.К. Бабенко, В.С. Стародубцев**
ОЦЕНКА ВРЕМЕНИ ВЫПОЛНЕНИЯ ОПЕРАЦИЙ ШИФРОВАНИЯ, РАСШИФРОВАНИЯ, ГОМОМОРФНЫХ ВЫЧИСЛЕНИЙ С ИСПОЛЬЗОВАНИЕМ КРИПТОСИСТЕМЫ ДОМИНГО-ФЕРРЕРА ..... 6

**А.А. Кабанов, В.А. Крамарь, К.В. Дементьев**
АЛГОРИТМ ОПТИМАЛЬНОГО КОМПЛЕКСИРОВАНИЯ ОЦЕНКИ СОСТОЯНИЙ В ДИСКРЕТНО-НЕПРЕРЫВНЫХ СИСТЕМАХ АНПА ..... 16

**В.В. Ковалев**
АЛГОРИТМ ПРЕДВАРИТЕЛЬНОЙ ОБРАБОТКИ ИЗОБРАЖЕНИЙ ДЛЯ СНИЖЕНИЯ ВЕРОЯТНОСТИ ПЕРЕОБУЧЕНИЯ СВЁРТОЧНЫХ НЕЙРОННЫХ СЕТЕЙ НА НЕЙРОННОМ УСКОРИТЕЛЕ ..... 29

**И.В. Котенко, М.В. Мельник**
ПРИМЕНЕНИЕ ГИБРИДНОЙ НЕЙРОННОЙ СЕТИ AE-LSTM ДЛЯ ОБНАРУЖЕНИЯ АНОМАЛИЙ В КОНТЕЙНЕРНЫХ СИСТЕМАХ ..... 38

**И.И. Левин, Е.А. Дудников**
СТРУКТУРНАЯ МОДИФИКАЦИЯ МЕТОДА ХАФФМАНА ДЛЯ СЖАТИЯ ПЛОТНЫХ ПОТОКОВ ДАННЫХ БЕЗ ПОТЕРЬ НА РВС ..... 48

**С.Ю. Мельников, Р.В. Мещеряков, В.А. Пересыпкин**
НЕКОТОРЫЕ АСПЕКТЫ ПРИМЕНЕНИЯ ТЕХНОЛОГИЙ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В ЗАДАЧАХ ЗАЩИТЫ ИНФОРМАЦИИ (ОБЗОР) ..... 58

**Е.А. Титенко, Э.И. Ватутин, М.А. Титенко, Э.В. Мельник, А.П. Локтионов**
АППАРАТНО-ОРИЕНТИРОВАННЫЙ МЕТОД УСКОРЕННОГО ПОИСКА ВХОЖДЕНИЙ ОБРАЗЦА НА ОСНОВЕ СТРУКТУРНО-ПРОЦЕДУРНЫХ ВЫЧИСЛЕНИЙ ..... 68

**И.В. Машкина, А.М. Уразаева**
МЕТОД РАЗРАБОТКИ БАЗЫ ЗНАНИЙ СЦЕНАРИЕВ УГРОЗ ДЛЯ СИСТЕМЫ РЕАГИРОВАНИЯ НА ИНЦИДЕНТЫ (IRP) ..... 79

**Е.М. Герасименко, Ю.А. Кравченко, Д.А. Шаненко**
АЛГОРИТМ ПОИСКА И ПРИОБРЕТЕНИЯ ЗНАНИЙ НА ОСНОВЕ ТЕХНОЛОГИЙ ОБРАБОТКИ И АНАЛИЗА ТЕКСТОВ НА ЕСТЕСТВЕННОМ ЯЗЫКЕ ..... 88

**С.В. Поликарпов, В.А. Прудников, К.Е. Румянцев**
СИНТЕЗ ПСЕВДО-ДИНАМИЧЕСКИХ ФУНКЦИЙ PD-sbox-ARX-32 ..... 102

**А.А. Александров, Г.С. Мизюков, М.А. Бутакова**
ОЦЕНКА КАЧЕСТВА СЛИЯНИЯ ИЗОБРАЖЕНИЙ С ИСПОЛЬЗОВАНИЕМ ЭНТРОПИИ ШЕННОНА И КОЭФФИЦИЕНТА ПОЛЕЗНОЙ ИНФОРМАЦИИ ХАРТЛИ ..... 119

**О.С. Малютин, Р.Ш. Хабибулин**
МЕТОДИКА ОПРЕДЕЛЕНИЯ ЧАСТОТЫ ВОЗНИКНОВЕНИЯ ПОЖАРОВ В ЗДАНИЯХ НА ОСНОВЕ МЕТОДОВ ОЦЕНКИ ПЛОТНОСТИ И ИМИТАЦИИ ОТЖИГА ..... 131

**Ж. Мохаммад**
ИЗВЛЕЧЕНИЕ КЛЮЧЕВЫХ ФРАЗ НА ОСНОВЕ БОЛЬШИХ ЯЗЫКОВЫХ МОДЕЛЕЙ ..... 143

### РАЗДЕЛ II. АНАЛИЗ ДАННЫХ И МОДЕЛИРОВАНИЕ

**В.И. Волощук, А. Гарягдыев, М.А. Козловская, Я.Э. Мельник, А.Н. Самойлов**
ПОДХОД К ПОСТРОЕНИЮ АДАПТИВНЫХ СИСТЕМ УЧЕТА ОБЪЕКТОВ С ИСПОЛЬЗОВАНИЕМ МЕТОДОВ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ..... 152

**В.О. Малявина, Е.А. Маро**
МОДЕЛИРОВАНИЕ УТЕЧЕК ПО ПОБОЧНЫМ КАНАЛАМ ДЛЯ КРИПТОГРАФИЧЕСКОГО АЛГОРИТМОВ «МАГМА» И «КУЗНЕЧИК» НА ОСНОВЕ ЭМУЛЯТОРА ELMO ..... 163

**Д.А. Сорокин, А.В. Касаркин**
ОБЗОР МОДЕЛЕЙ КОММУТАЦИОННЫХ ПОДСИСТЕМ ЦИФРОВЫХ ФОТОННЫХ ВЫЧИСЛИТЕЛЬНЫХ УСТРОЙСТВ ..... 173

**М. Пленингер, С.В. Балакирев, М.С. Солодовник**
МОДЕЛИРОВАНИЕ РАСПРЕДЕЛЕНИЯ НАПРЯЖЕННОСТИ ЭЛЕКТРИЧЕСКОГО ПОЛЯ В ПОЛНОСТЬЮ ОПТИЧЕСКОМ ЛОГИЧЕСКОМ КОМПАРАТОРЕ НА ОСНОВЕ ФОТОННОГО КРИСТАЛЛА GaAs ..... 185

**А.Н. Самойлов, Н.Е. Сергеев, С.М. Гушанский, В.С. Потапов**
РАЗРАБОТКА И ИССЛЕДОВАНИЕ КВАНТОВОЙ ГРАФОВОЙ МОДЕЛИ ДЛЯ СЖАТИЯ И РЕКОНСТРУКЦИИ ИЗОБРАЖЕНИЙ ..... 194

**РАЗДЕЛ III. ЭЛЕКТРОНИКА, ПРИБОРОСТРОЕНИЕ И РАДИОТЕХНИКА**

**И.Н. Бобков, Ю.В. Юханов**
ДВУХПОЛЯРИЗАЦИОННАЯ АНТЕННАЯ РЕШЕТКА ВИВАЛЬДИ С УМЕНЬШЕННОЙ ВЫСОТОЙ ПРОФИЛЯ ..... 205

**А.А. Жук**
СХЕМОТЕХНИЧЕСКИЕ МЕТОДЫ ПОВЫШЕНИЯ ЭФФЕКТИВНОСТИ ВЫХОДНЫХ КАСКАДОВ АРСЕНИД-ГАЛЛИЕВЫХ ОПЕРАЦИОННЫХ УСИЛИТЕЛЕЙ НОВОГО ПОКОЛЕНИЯ ..... 214

**Р.Э. Косак**
АНТЕННАЯ РЕШЕТКА КОМПАКТНЫХ ИЗЛУЧАТЕЛЕЙ ВИВАЛЬДИ С ЭЛЛИПТИЧЕСКИМИ ВЫРЕЗАМИ НА КРОМКЕ ..... 232

**И.И. Левин, Д.С. Буряков**
НЕКОТОРЫЕ МЕТОДЫ СИНХРОНИЗАЦИИ ИНФОРМАЦИОННЫХ ПОТОКОВ В СИСТЕМАХ ЦИФРОВОЙ ОБРАБОТКИ СИГНАЛОВ ..... 243

**А.П. Плёнкин**
СПОСОБ ОБНАРУЖЕНИЯ ОПТИЧЕСКОГО СИГНАЛА В КВАНТОВЫХ СЕТЯХ ..... 254

**А.В. Геворкян, В.С. Савостин**
СВЕРХШИРОКОПОЛОСНЫЕ РЕШЁТКИ АНТЕНН ВИВАЛЬДИ С ТЕМ-РУПОРОМ ..... 260

**И.А. Калмыков, И.Д. Ефременков, Д.В. Духовный**
ПОМЕХОУСТОЙЧИВЫЙ ПРОТОКОЛ ОПОЗНАВАНИЯ НИЗКООРБИТАЛЬНОГО СПУТНИКА-РЕТРАНСЛЯТОРА ..... 271

**РАЗДЕЛ IV. СООБЩЕНИЕ ОБ ОТЗЫВЕ ПУБЛИКАЦИИ** ..... 283

## CONTENT

### SECTION I. INFORMATION PROCESSING ALGORITHMS

**L.K. Babenko, V.S. Starodubtsev**
ESTIMATION OF THE EXECUTION TIME OF ENCRYPTION, DECRYPTION, AND HOMOMORPHIC CALCULATIONS USING THE DOMINGO-FERRER CRYPTOSYSTEM ..... 6

**A.A. Kabanov, V.A. Kramar, K.V. Dementiev**
THE OPTIMAL STATE ESTIMATION FUSION ALGORITHM IN DISCRETE-CONTINUOUS AUV SYSTEMS ..... 16

**V.V. Kovalev**
IMAGE PREPROCESSING ALGORITHM TO REDUCE THE PROBABILITY OF OVERFITTING OF CONVOLUTIONAL NEURAL NETWORKS ON A NEURAL ACCELERATOR ..... 29

**I.V. Kotenko, M.V. Melnik**
APPLICATION OF A HYBRID NEURAL NETWORK AE-LSTM FOR ANOMALIES DETECTION IN CONTAINER SYSTEMS..... 38

**I.I. Levin, E.A. Dudnikov**
STRUCTURAL MODIFICATION OF THE HUFFMAN METHOD FOR COMPRESSION OF DENSE DATA STREAMS WITHOUT LOSS ON A RCS..... 48

**S.Yu. Melnikov, R.V. Meshcheryakov, V.A. Peresypkin**
SOME ASPECTS OF APPLICATION OF ARTIFICIAL INTELLIGENCE TECHNOLOGIES IN INFORMATION SECURITY (REVIEW) ..... 58

**E.A. Titenko, E.I. Vatutin, M.A. Titenko, E.V. Melnik, A.P. Loktionov**
A HARDWARE-ORIENTED METHOD OF ACCELERATED SEARCH BY TEMPLATE BASED ON STRUCTURAL-PROCEDURAL COMPUTING ..... 69

**I.V. Mashkina, A.M. Urazaeva**
METHOD OF DEVELOPMENT OF THREAT SCENARIOS KNOWLEDGE BASE FOR INCIDENT RESPONSE PLATFORM (IRP)..... 79

**E.M. Gerasimenko, Yu.A. Kravchenko, D.A. Shanenko**
ALGORITHM FOR SEARCHING AND ACQUISITION OF KNOWLEDGE BASED ON TECHNOLOGIES FOR PROCESSING AND ANALYZING TEXTS IN NATURAL LANGUAGE..... 89

**S.V. Polikarpov, V.A. Prudnikov, K.E. Rumyantsev**
SYNTHESIS OF PSEUDO-DYNAMIC FUNCTIONS PD-sbox-ARX-32..... 103

**A.A. Alexandrov, G.S. Miziukov, M.A. Butakova**
IMAGE FUSION QUALITY ASSESSMENT USING SHANNON ENTROPY AND HARTLEY USEFUL INFORMATION COEFFICIENT..... 119

**O.S. Malyutin , R.Sh. Khabibulin**
BUILDINGS FIRES FREQUENCY DETERMINING METHODOLOGY BASED ON DENSITY ESTIMATION AND SIMULATED ANNEALING METHODS ..... 132

**J. Mohammad**
KEYPHRASE EXTRACTION BASED ON LARGE LANGUAGE MODELS ..... 143

### SECTION II. DATA ANALYSIS AND MODELING

**V.I. Voloshchuk, A. Garyagdiyev, M.A. Kozlovskaya, Y.E. Melnik, A.N. Samoylov**
AN APPROACH TO BUILDING ADAPTIVE OBJECT ACCOUNTING SYSTEMS USING ARTIFICIAL INTELLIGENCE METHODS ..... 152

**V.O. Malyavina, E.A. Maro**
MODELING SIDE-CHANNEL LEAKAGES FOR THE CRYPTOGRAPHIC ALGORITHMS "MAGMA" AND "KUZNACHIK" BASED ON THE ELMO EMULATOR..... 163

**D.A. Sorokin, A.V. Kasarkin**
OVERVIEW OF SWITCHING SUBSYSTEM MODELS FOR DIGITAL PHOTONIC COMPUTING DEVICES ..... 174

**M. Pleninger, S.V. Balakirev, M.S. Solodovnik**
SIMULATION OF THE ELECTRIC FIELD STRENGTH DISTRIBUTION IN AN ALL-OPTICAL LOGIC COMPARATOR BASED ON THE GaAs PHOTONIC CRYSTAL..... 185

**A.N. Samoilov, N.E. Sergeev, S.M. Gushanskiy, V.S. Potapov**
DEVELOPMENT AND RESEARCH OF A QUANTUM GRAPH MODEL FOR IMAGE COMPRESSION AND RECONSTRUCTION ..... 195

**SECTION III. ELECTRONICS, INSTRUMENTATION AND RADIO ENGINEERING**

**I.N. Bobkov, Y.V. Yukhanov**
A DUAL-POLARIZED TAPERED SLOT ANTENNA ARRAY WITH REDUCED PROFILE HEIGHT..... 205

**A.A. Zhuk**
HIGH-SPEED OUTPUT STAGES OF OPERATIONAL AMPLIFIERS WITH DIFFERENCING CIRCUIT CORRECTION OF TRANSITION PROCESS ..... 215

**R.E. Kosak**
ANTENNA ARRAY OF COMPACT VIVALDI RADIATORS WITH ELLIPTICALL SHAPE CUTOUTS ON THEIR OUTER EDGE..... 233

**I.I. Levin, D.S. Buryakov**
SOME METHODS FOR DATA FLOW SYNCHRONIZATION IN DIGITAL SIGNAL PROCESSING SYSTEMS ..... 243

**A.P. Pljonkin**
METHOD FOR DETECTING OPTICAL SIGNAL IN QUANTUM NETWORKS ..... 255

**A.V. Gevorkyan, V.S. Savostin**
ULTRA-WIDEBAND VIVALDI ANTENNA ARRAYS WITH TEM HORN ..... 261

**I.A. Kalmykov, I.D. Efremenkov, D.V. Dukhovnyj**
NOISE-RESISTANT LOW-ORBIT REPEATER SATELLITE IDENTIFICATION PROTOCOL..... 271

**SECTION IV. REPORT OF RETRACTION**..... 283

# Раздел I. Алгоритмы обработки информации

УДК 004.056.55 DOI 10.18522/2311-3103-2024-5-6-15

**Л.К. Бабенко, В.С. Стародубцев**

## ОЦЕНКА ВРЕМЕНИ ВЫПОЛНЕНИЯ ОПЕРАЦИЙ ШИФРОВАНИЯ, РАСШИФРОВАНИЯ, ГОМОМОРФНЫХ ВЫЧИСЛЕНИЙ С ИСПОЛЬЗОВАНИЕМ КРИПТОСИСТЕМЫ ДОМИНГО-ФЕРРЕРА

*Рассматривается симметричная вероятностная гомоморфная криптосистема Доминго-Феррера, основанная на задаче факторизации чисел. В настоящее время актуальны гомоморфные криптосистемы двух типов: типа Джентри и основанные на задаче факторизации чисел. Отличительной особенностью последних по сравнению с криптосистемами типа Джентри является меньшая трудоёмкость выполнения гомоморфных операций, что значительно расширяет область их применения на практике. Однако, поскольку гомоморфные криптосистемы, основанные на задаче факторизации чисел, не получили широкого распространения и не были в достаточной мере проанализированы, в отличие от криптосистем типа Джентри, требуется их тщательное всестороннее исследование. Для рассматриваемой симметричной гомоморфной криптосистемы Доминго-Феррера приводятся описания операций генерации ключа, шифрования, расшифрования и выполнения гомоморфных вычислений. Для операций шифрования, расшифрования и выполнения гомоморфных вычислений приводится оценка сложности, выраженная в количестве базовых математических операций, а также графики, иллюстрирующие зависимости количества операций от выбранных параметров криптосистемы. Целью исследования является оценка сложности выполнения процессов шифрования, расшифрования и выполнения гомоморфных вычислений симметричной вероятностной гомоморфной криптосистемой Доминго-Феррера, основанной на задаче факторизации чисел. Основным результатом настоящей работы является оценка сложности и определение наиболее трудоёмких этапов шифрования, расшифрования и выполнения гомоморфных вычислений с помощью шифра Доминго-Феррера, подтвержденных рядом экспериментальных исследований. Проведенное исследование представляет собой важный шаг в развитии криптографической системы Доминго-Феррера, основанной на задаче факторизации чисел, имеет практическую значимость реализации алгоритмов с возможностью определения временных затрат шифрования, расшифрования и выполнения гомоморфных вычислений. Полученные результаты могут быть использованы исследователями и программистами при разработке реализаций криптосистемы Доминго-Феррера на языках программирования.*

*Информационная безопасность; конфиденциальная информация; гомоморфное шифрование; криптосистема Доминго-Феррера; криптоанализ; оценка сложности алгоритма шифрования.*

**L.K. Babenko, V.S. Starodubtsev**

## ESTIMATION OF THE EXECUTION TIME OF ENCRYPTION, DECRYPTION, AND HOMOMORPHIC CALCULATIONS USING THE DOMINGO-FERRER CRYPTOSYSTEM

*This article considers a symmetric probabilistic homomorphic Domingo-Ferrer cryptosystem based on the problem of number factorization. Currently, homomorphic cryptosystems of two types are relevant: the Gentry type and those based on the problem of factorization of numbers. A distinctive feature of the latter, in comparison with Gentry-type cryptosystems, is the lower complexity of performing homomorphic operations, which significantly expands the scope of their application in practice. However, since homomorphic cryptosystems based on the number factorization problem have not been widely used and have not been sufficiently analyzed, unlike Gentry-type cryptosystems, their thorough comprehensive study is required. For the considered symmetric homomorphic Domingo-Ferrer cryptosystem, descriptions of*

*key generation, encryption, decryption, and homomorphic computing operations are given. For encryption, decryption, and homomorphic computing operations, a complexity estimate is given, expressed in the number of basic mathematical operations, as well as graphs illustrating the dependence of the number of operations on the selected parameters of the cryptosystem. The aim of the study is to assess the complexity of performing encryption, decryption and homomorphic calculations by a symmetric probabilistic homomorphic Domingo-Ferrer cryptosystem based on the number factorization problem. The main result of this work is an assessment of the complexity and determination of the most time-consuming stages of encryption, decryption and performing homomorphic calculations using the Domingo-Ferrer cipher, confirmed by a number of experimental studies. The conducted research represents an important step in the development of the Domingo-Ferrer cryptographic system based on the problem of factorization of numbers and has the practical significance of implementing algorithms with the ability to determine the time costs of encryption, decryption and performing homomorphic calculations. The results obtained can be used by researchers and programmers in the development of implementations of the Domingo-Ferrer cryptosystem in programming languages.*

*Information security; confidential information; homomorphic encryption; Domingo-Ferrer cryptosystem; cryptanalysis; evaluation of the complexity of the encryption algorithm.*

**Введение.** Гомоморфное шифрование является одной из важных техник в области криптографии, которая предоставляет возможность выполнять операции с зашифрованными данными без необходимости расшифровывать их. Это особенно полезно в контексте облачных вычислений и обработки данных, где часто возникает необходимость в анализе и использовании конфиденциальной информации.

Гомоморфное шифрование – современная техника, позволяющая значительно расширить область применения криптографии для защиты информации. Главной отличительной чертой гомоморфного шифрования является возможность выполнять различные операции над данными в зашифрованном виде без необходимости их предварительного расшифрования. Данное свойство крайне важно при использовании технологий облачных вычислений, поскольку позволяет применять данную технологию для обработки конфиденциальной информации.

В настоящее время существуют два основных типа гомоморфных криптосистем: типа Джентри и основанные на задаче факторизации чисел. Криптосистемы типа Джентри [1–4] получили широкое распространение и были всесторонне исследованы, для данных криптосистем доказана их высокая криптографическая стойкость. Однако гомоморфные операции в криптосистемах типа Джентри имеют высокую вычислительную сложность, ввиду чего выполняются крайне медленно и значительно сужают область их применения на практике. Криптосистемы, основанные на задаче факторизации чисел, обладают меньшей трудоёмкостью гомоморфных операций по сравнению с криптосистемами типа Джентри и имеют больший потенциал для применения на практике, однако на данный момент не обрели популярности и не были в достаточной мере проанализированы, и, как следствие, не была проведена оценка трудоёмкости. Поэтому оценка времени выполнения гомоморфных вычислений с помощью шифров, основанных на задаче факторизации чисел, является актуальной задачей.

**Описание криптосистемы Доминго-Феррера.** Эта криптосистема была создана в 1996 году и поддерживает гомоморфное сложение, вычитание и умножение [5–7]. Она является симметричной, то есть для шифрования и расшифрования используется один и тот же ключ [8]. Количество последовательных гомоморфных операций в данной системе не ограничено, однако размеры итоговых шифртекстов увеличиваются; например, при умножении размер шифртекста возрастает экспоненциально [9, 10]. Для инициализации криптосистемы Доминго-Феррера используется следующий набор чисел:

1) $p$ и $q$ – большие простые числа;

2) $n = p \times q$ – труднофакторизуемое число;

3) $d$ – степень полиномов представления шифртекстов.

Алгоритмы генерации ключа, шифрования и расшифрования криптосистемой Доминго-Феррера приведены на рис. 1.


*Рис. 1. Описание операций шифра Доминго-Феррера*

Ключ в криптосистеме Доминго-Феррера представляется двумя числами – $r_p$ и $r_q$, которые случайно выбираются из мультипликативных групп с соответствующим модулем: формулы (1) и (2) соответственно.

$$r_p \xleftarrow{\$} Z_p^*, \tag{1}$$

где $Z_p^*$ – мультипликативная группа по модулю $p$, $\xleftarrow{\$}$ – операция выбора случайного элемента.

$$r_q \xleftarrow{\$} Z_q^*, \tag{2}$$

где $Z_q^*$ – мультипликативная группа по модулю $q$, $\xleftarrow{\$}$ – операция выбора случайного элемента.

**Количество операций выбора случайного элемента при шифровании.** Для выполнения шифрования в криптосистеме Доминго-Феррера необходимо убедиться, что блок шифруемого сообщения $m \in Z_n$. Затем генерируется ряд случайных значений $a_1, \ldots, a_d$.Числам $a_2, \ldots, a_{d-1}$ присваивается значение по формуле (3).

$$a_i \xleftarrow{\$} Z_n, \tag{3}$$

где $i$ – индекс числа в диапазоне [2; d-1], $Z_n$ – кольцо по модулю $n$, $\xleftarrow{\$}$ – операция выбора случайного элемента.

Числу $a_d$ аналогичным образом присваивается случайное значение из $Z_n \setminus \{0\}$ (любое ненулевое значение из $Z_n$) по формуле (4).

$$a_d \xleftarrow{\$} Z_n \setminus \{0\}. \tag{4}$$

Таким образом, из представленных формул (3) и (4) видно, что количество операций выбора случайного элемента зависит от $d$. Для ряда $a_2 \ldots a_d$ необходимо выбрать $d - 1$ случайных элементов.

**Количество операций сложения и вычитания при шифровании.** Когда определен набор случайных чисел $a_2 \ldots a_d$, значение $a_1$вычисляется по формуле (5).

$$a_1 = m - \left(\sum_{i=2}^{d} a_i\right) mod\ n, \tag{5}$$

где $m$ – блок открытого текста, $n$ – модуль, определенный в параметрах схемы шифрования.

Формула (5) содержит сложение сгенерированных случайных чисел из набора $a_2 \ldots a_d$. Всего в формуле (5) присутствует $d - 1$ слагаемых, следовательно для их сложения необходимо выполнить $d - 2$ операций сложения. Также в формуле (5) используется одна операция вычитания полученной суммы $\sum_{i=2}^{d} a_i$ из значения блока открытого текста $m$.

Сформированный набор чисел $a_1 \dots a_d$ представляет собой закодированный открытый текст и представляется в виде полинома по формуле (6). На данном этапе никаких математических операций не выполняется.

$$a(x) = a_d x^d + \cdots + a_1 x. \quad (6)$$

**Количество операций умножения при шифровании.** Когда открытый текст закодирован, происходит его шифрование с помощью составного ключа $(r_p, r_q)$. Полином $\pi(x)$ формируется путём шифрования $a(x)$ на первой части ключа $r_p$ по формуле (7).

$$\pi(x) = \left(a_d \times r_p^d x^d + \cdots + a_1 \times r_p x\right) mod\ p. \quad (7)$$

Полином $\rho(x)$ формируется аналогичным образом – путём шифрования $a(x)$ на второй части ключа $r_q$ по формуле (8).

$$\rho(x) = \left(a_d \times r_q^d x^d + \cdots + a_1 \times r_q x\right) mod\ q. \quad (8)$$

Сформированная пара $(\pi(x), \rho(x))$ – зашифрованное сообщение $m$.

Из формул (7) и (8) следует, что для формирования каждого из полиномов шифртекста $(\pi(x), \rho(x))$ происходит умножение каждого коэффициента полинома $a(x)$ на соответствующую часть ключа, возведенную в степень полинома, перед которой данный коэффициент установлен [11–14]. Следовательно, количество операций умножения при формировании полиномов шифртекста $(\pi(x), \rho(x))$ определяется по формуле (9).

$$MulCount = 2\sum_{i=1}^{d} i, \quad (9)$$

где $d$ – степень полинома представления шифртекста.

На рис. 2 приводится график зависимости количества операций умножения при шифровании криптосистемой Доминго-Феррера от выбранного значения $d$ (степени полинома представления шифртекстов).


*Рис. 2. Зависимость количества операций умножения от степени полинома представления шифртекста при шифровании*

Как видно из рис. 2, график имеет вид, близкий к функции $y = x^2$, что показывает, что сложность операции шифрования по умножению является квадратичной.

**Оценка количества операций получения остатка от деления при шифровании.** В шифре Доминго-Феррера операция получения остатка от деления при шифровании используется на заключительном этапе формирования полинома $a(x)$ – при вычислении числа $a_1$ (формула (3)), а также в формулах (7) и (8) на этапе вычисления значений шифртекста $(\pi(x), \rho(x))$. При формировании числа $a_1$ операции умножения не выполняются, поэтому достаточно только одной операции $mod\ n$. При вычислении полиномов шифртекста $(\pi(x), \rho(x))$ выполняются операции умножения, поэтому необходимо приведение по модулю (для полинома $\pi(x)$ по модулю $p$, для полинома $\rho(x)$ – по модулю $q$) после каждой операции умножения и их общее количество равно количеству операций умножения – $2\sum_{i=1}^{d} i$. Таким образом, при шифровании в криптосистеме Доминго-Феррера общее количество операций получения остатка от деления $ModCount$ вычисляется по формуле (10).

$$ModCount = 2\sum_{i=1}^{d} i + 1, \quad (10)$$

где $d$ – степень полинома представления шифртекста.

Таким образом, исходя из результатов оценки сложности процесса шифрования криптографической системы Доминго-Феррера видно, что в данном процессе выполняются:

- $d - 1$ операций выбора случайного элемента;
- $d - 2$ операций сложения;
- 1 операция вычитания;
- $2\sum_{i=1}^{d} i$ умножений;
- $2\sum_{i=1}^{d} i + 1$ операций получения остатка от деления.

Для подтверждения полученных результатов разработана реализация шифра Доминго-Феррера на языке C++ и проведены экспериментальные исследования.

При выбранном значении степени полинома представления шифртекста $d = 2048$ оценочное число процессорных тактов составило 17037369, согласно [15]. На одном ядре процессора AMD Ryzen 5 3500U (2.1 ГГц) время шифрования одного блока – 34 мс.

**Применение Китайской теоремы об остатках при расшифровании.** В криптосистеме Доминго-Феррера на заключительном этапе расшифрования, в отличие от шифрования, происходит поиск открытого текста, как решение СЛАУ по модулям $p$ и $q$ с применением Китайской теоремы об остатках [16–18] по формуле (11).

$$m = CRT(\{M_p, M_q\}, \{p, q\}), \quad (11)$$

где $CRT$ – функция, использующая Китайскую теорему об остатках для поиска значения открытого текста $m$.

Поиск решения системы линейных сравнений по модулю с применением доказательства Китайской теоремы об остатках [19] приведен в формуле (12).

$$x = \sum_{i=1}^{n} a_i \times M_i \times N_i \ mod\ M, \quad (12)$$

где $a_i$ – целое число, $M_i = \frac{M}{p_i}$, $N_i = M_i^{-1}$, $M = p_1 \times p_2 \times \ldots \times p_n$.

Как видно из формулы (12), $N_i$ определяется, как мультипликативное обратное $M_i$, которое вычисляется с помощью расширенного алгоритма Евклида [20]. Исходя из того, что на вход расширенного алгоритма Евклида подаются пары значений $(p, q)$ и $(q, p)$ количество выполняемых математических операций в процессе расшифрования не зависит от степени полинома представления шифртекста $d$, но зависит от выбранных параметров $p$ и $q$. Поэтому сложность расширенного алгоритма Евклида рассматривается отдельно.

**Оценка количества шагов расширенного алгоритма Евклида для поиска мультипликативного обратного.** Описание расширенного алгоритма Евклида приведено на рис. 3. На вход алгоритма подается число $a$, для которого необходимо найти мультипликативное обратное в группе по модулю $p$.

$(x_1, x_2, x_3) = (1, 0, p)$;
$(y_1, y_2, y_3) = (0, 1, a)$;
**Пока** $y_3 \neq 0$ & $y_3 \neq 1$:
    **Если** $y_3 = 0$, то:
$a^{-1} \nexists$;
конец.
    **Если** $y_3 = 1$, то:
        $a^{-1} = y_2$;
        конец.
    $Q = \frac{x_3}{y_3}$;
    $(T_1, T_2, T_3) = (x_1 - Q \cdot y_1; x_2 - Q \cdot y_2; x_3 - Q \cdot y_3)$;
    $x = y; y = T$;

*Рис. 3. Расширенный алгоритм Евклида*

Для оценки количества шагов расширенного алгоритма Евклида на различных наборах данных, значения параметров $p$ и $q$ принимали значения простых чисел из диапазона (10; 10000). Результаты оценки приводятся на рис. 4. По оси абсцисс приведены значения числа $q$, по оси ординат – значения числа $p$. Чем светлее пиксель – тем больше шагов требуется для поиска мультипликативных обратных с помощью расширенного алгоритма Евклида.

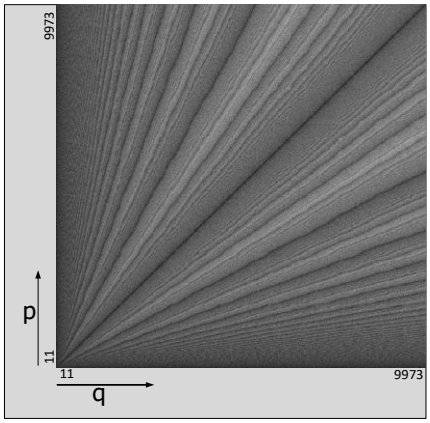



*Рис. 4. Количество шагов расширенного алгоритма Евклида в зависимости от пары входных чисел (p, q)*

В результате проведенного исследования минимальное число шагов расширенного алгоритма Евклида при расшифровании в криптосистеме Доминго-Феррера составило 2 (случаи, когда $p = q$), максимальное – 33 (например, при значениях $p = 3571, q = 9349$). В среднем криптосистеме Доминго-Феррера требовалось 14 шагов расширенного алгоритма Евклида для выполнения расшифрования. Как видно из рис. 3, на каждом шаге расширенного алгоритма Евклида [20] выполняется 1 деление, 3 вычитания и 3 умножения, то в среднем для выполнения расшифрования криптосистеме Доминго-Феррера требовалось 14 делений, 42 вычитания и 42 умножения.

**Количество операций умножения при расшифровании.** В шифре Доминго-Феррера на начальном этапе расшифрования $\left(\pi(x), \rho(x)\right)$ снимаются значения ключа $(r_p, r_q)$ путём умножения на их мультипликативные обратные $(r_p^{-1}, r_q^{-1})$, возведенные в соответствующие степени $A_p(x)$ и $A_q(x)$ по формулам (13) и (14) соответственно.

$$A_p(x) = \left(b_d \times (r_p^{-1})^d x^d + \cdots + b_1 \times (r_p^{-1})\, x\right) mod\ p, \qquad (13)$$

где b – коэффициенты полинома $\pi(x)$ шифртекста.

$$A_q(x) = \left(b_d \times (r_q^{-1})^d x^d + \cdots + b_1 \times (r_q^{-1})\, x\right) mod\ q, \qquad (14)$$

где b – коэффициенты полинома $\rho(x)$ шифртекста.

Так как в симметричной криптосистеме Доминго-Феррера ключ не изменяется, для его составляющих $\left(r_p, r_q\right)$ нет необходимости при каждом расшифровании вычислять мультипликативные обратные, поэтому принимается, что мультипликативные обратные $\left(r_p^{-1}, r_q^{-1}\right)$ вычисляются в процессе генерации ключа и хранятся в памяти.

Из формул (13) и (14) следует, что для формирования каждого из полиномов $\left(A_p(x)\,, A_q(x)\right)$ происходит умножение каждого коэффициента полиномов шифртекста $\left(\pi(x), \rho(x)\right)$ на мультипликативное обратное соответствующей части ключа, возведенное в степень полинома, перед которой данный коэффициент установлен [11–14]. Следовательно, количество операций умножения при формировании полиномов $\left(A_p(x)\,, A_q(x)\right)$ равно количеству проводимых умножений в операции шифрования.

Однако, в шифре Доминго-Феррера в отличие от шифрования, в операции расшифрования умножение также применяется в расширенном алгоритме Евклида (для значений $p$ и $q$ до 10000 – в среднем 42 раза) и в Китайской теореме об остатках 5 раз.

Следовательно, общее количество умножений $MulCount$, необходимых криптосистеме Доминго-Феррера с параметрами $p$ и $q$ до 10000 для расшифрования шифртекста определяется по формуле (15).

$$MulCount = 2\sum_{i=1}^{d} i + 5 + 42, \quad (15)$$

где $d$ – степень полинома представления шифртекста.

На рис. 5 приводится график зависимости количества операций умножения при расшифровании криптосистемой Доминго-Феррера от выбранного значения $d$ (степени полинома представления шифртекстов).


*Рис. 5. Зависимость количества операций умножения от степени полинома представления шифртекста при расшифровании*

Как видно из рис. 5, график имеет вид, близкий к функции $y = x^2$, что показывает, что сложность операции расшифрования по умножению является квадратичной.

**Количество операций сложения и вычитания при расшифровании.** После снятия значений ключа $(r_p, r_q)$ для полиномов $\left(A_p(x), A_q(x)\right)$ по формулам (16) и (17) вычисляются суммы $M_p$ и $M_p$ соответственно.

$$M_p = \sum_{i=1}^{d} a_i \ mod\ p, \quad 16)$$

где $a_i$ – коэффициенты полинома $A_p(x)$.

$$M_q = \sum_{i=1}^{d} b_i \ mod\ q, \quad (17)$$

где $b_i$ – коэффициенты полинома $A_q(x)$.

Формулы (16) и (17) содержат сложение коэффициентов полиномов $\left(A_p(x), A_q(x)\right)$. Всего в формулах (16) и (17) присутствует по $d$ слагаемых, следовательно для их сложения необходимо выполнить $2(d-1)$ операций сложения.

Важно отметить, что в шифре Доминго-Феррера в отличие от шифрования, в операции расшифрования операции сложения и вычитания также применяются в расширенном алгоритме Евклида (для значений $p$ и $q$ до 10000 – в среднем 42 вычитания) и в Китайской теореме об остатках (2 сложения). Следовательно, в процессе расшифрования требуется выполнить $2(d-1)+2$ операций сложения и 42 операции вычитания.

**Количество операций получения остатка от деления при расшифровании.** Формулы (13) и (14) показывают, что для формирования полиномов $A_p(x)$ и $A_q(x)$, а также подсчёта сумм их коэффициентов $M_p$ и $M_q$ по формулам (16) и (17) соответственно требуются операции получения остатков от деления. Помимо этого, на заключительном

этапе расшифрования выполняется операция получения остатка от деления на $n$ в Китайской теореме об остатках, что приводит к общему количеству получения остатка от деления $2\sum_{i=1}^{d} i + 3$.

Таким образом, исходя из результатов оценки сложности процесса расшифрования криптографической системы Доминго-Феррера видно, что в данном процессе выполняются:

- $2(d-1)+2$ операций сложения;
- 42 операции вычитания;
- $2\sum_{i=1}^{d} i + 5 + 42$ умножений;
- $2\sum_{i=1}^{d} i + 3$ операций получения остатка от деления.

Для подтверждения полученных результатов разработана реализация шифра Доминго-Феррера на языке C++ и проведены экспериментальные исследования.

При выбранном значении степени полинома представления шифртекста $d = 2048$ оценочное число процессорных тактов составило 17036373, согласно [15]. На одном ядре процессора AMD Ryzen 5 3500U (2.1 ГГц) время расшифрования одного блока – 32 мс.

После вычисления сумм $M_p$ и $M_q$ открытый текст исходного сообщения рассчитывается с помощью Китайской теоремы об остатках по формуле (18).

$$m = CRT\big(\{M_p, M_q\}, \{p, q\}\big), \tag{18}$$

где $CRT()$ – функция, использующая Китайскую теорему об остатках для поиска значения открытого текста $m$.

**Оценка времени выполнения гомоморфных операций.** Для выполнения операции (сложение, вычитание, умножение) над двумя шифртекстами в криптосистеме Доминго-Феррера нужно применить эту операцию к полиномам шифртекстов, как показано в формуле (19).

$$C_3 = \{(\pi(x)_1 \circ \pi(x)_2), (\rho(x)_1 \circ \rho(x)_2)\}, \tag{19}$$

где $C_3$ – результирующий шифртекст, $(\pi(x), \rho(x))_1$ – первый шифртекст, $(\pi(x), \rho(x))_2$ – второй шифртекст.

Важно, что выбранная математическая операция покоординатная, умножаются (складываются) коэффициенты полиномов только с одинаковой степенью.

Следовательно, выполнение гомоморфных операций в криптосистеме Доминго-Феррера обладает линейной сложностью $O(d)$ вне зависимости от того, какая именно гомоморфная операция выполняется.

**Выводы.** В данной работе проведена оценка сложности выполнения операций шифрования, расшифрования и гомоморфных вычислений симметричной вероятностной гомоморфной криптосистемой Доминго-Феррера, приведены формулы для расчёта количества математических операций в зависимости от выбранных параметров (степени полиномов представления шифртекста d, простых чисел p и q). Полученные оценки подтверждены экспериментальными исследованиями реализации шифра Доминго-Феррера на языке программирования C++. С использованием данной реализации время шифрования и расшифрования 1 блока размером 64 бита со значением модуля $n$ до $10^8$ и степенью полинома представления шифртекста $d = 2048$ на процессоре AMD Ryzen 5 3500U (2.1 ГГц) в однопоточном режиме оказалось примерно равным и составило 33 мс.

Таком образом, можно сделать вывод, что в криптографической системе Доминго-Феррера как при шифровании, так и при расшифровании наибольшее количество операций выполняется на этапах сопряжения с ключом, поскольку для каждого коэффициента полинома требуется возведение ключа в соответствующую степень. Это приводит к тому, что для шифрования или расшифрования требуется $2\sum_{i=1}^{d} i$ операций умножения, что является квадратичной сложностью алгоритма $O(d^2)$ и в контексте реализации на языках программирования и дальнейшего использования на практике требует оптимизации.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Fan J., Vercauteren F.* Somewhat practical fully homomorphic encryption // Cryptology ePrint Archive. – 2012.
2. *Gentry C., Sahai A., Waters B.* Homomorphic encryption from learning with errors: Conceptually-simpler, asymptotically-faster, attribute-based // Advances in Cryptology–CRYPTO 2013: 33rd Annual Cryptology Conference, Santa Barbara, CA, USA, August 18-22, 2013. Proceedings, Part I. – Springer Berlin Heidelberg, 2013. – P. 75-92.
3. *Brakerski Z.* Fully homomorphic encryption without modulus switching from classical GapSVP // Annual cryptology conference. – Berlin, Heidelberg: Springer Berlin Heidelberg, 2012. – P. 868-886.
4. *Brakerski Z., Gentry C., Vaikuntanathan V.* (Leveled) fully homomorphic encryption without bootstrapping // ACM Transactions on Computation Theory (TOCT). – 2014. – Vol. 6, No. 3. – P. 1-36.
5. *Domingo-Ferrer J.* A provably secure additive and multiplicative privacy homomorphism // International Conference on Information Security. – Berlin, Heidelberg: Springer Berlin Heidelberg, 2002. – P. 471-483.
6. *Hariss K., Noura H., Samhat A. E.* An efficient fully homomorphic symmetric encryption algorithm // Multimedia Tools and Applications. – 2020. – Vol. 79, No. 17. – P. 12139-12164.
7. *Wang H., Wang Z., Domingo-Ferrer J.* Anonymous and secure aggregation scheme in fog-based public cloud computing // Future Generation Computer Systems. – 2018. – Vol. 78. – P. 712-719.
8. *Maqsood F. et al.* Cryptography: a comparative analysis for modern techniques // International Journal of Advanced Computer Science and Applications. – 2017. – Vol. 8, No. 6.
9. *Трепачева А. В.* Улучшенная атака по известным открытым текстам на гомоморфную криптосистему Доминго-Феррера // Тр. Института системного программирования РАН. – 2014. – Т. 26, № 5. – С. 83-98.
10. *Alabdulatif A., Kaosar M.* Privacy preserving cloud computation using Domingo-Ferrer scheme // Journal of King Saud University-Computer and Information Sciences. – 2016. – Vol. 28, No. 1. – P. 27-36.
11. *Merkel W. et al.* Factorization of numbers with physical systems // Fortschritte der Physik: Progress of Physics. – 2006. – Vol. 54, No. 8-10. – P. 856-865.
12. *Lenstra A. K. et al.* The factorization of the ninth Fermat number // Mathematics of Computation. – 1993. – Vol. 61, No. 203. – P. 319-349.
13. *Яковлев В.А., Шемякин С.Н., Таров Е.В.* Использование метода Монтгомери в алгоритме быстрого возведения в степень // I-methods. – 2023. – Vol. 15, No. 1. – P. 6.
14. *Hossain M.A. et al.* Performance analysis of different cryptography algorithms // International Journal of Advanced Research in Computer Science and Software Engineering. – 2016. – Vol. 6, No. 3.
15. *Agner F. Optimizing software in C++:* An optimization guide for Windows, Linux and Mac platforms. – 2020.
16. *Pei D., Salomaa A., Ding C.* Chinese remainder theorem: applications in computing, coding, cryptography. – World Scientific, 1996.
17. *Schindler W.* A timing attack against RSA with the chinese remainder theorem // Cryptographic Hardware and Embedded Systems—CHES 2000: Second International Workshop Worcester, MA, USA, August 17–18, 2000 Proceedings 2. – Springer Berlin Heidelberg, 2000. – P. 109-124.
18. *Wang W., Xia X. G.* A closed-form robust Chinese remainder theorem and its performance analysis // IEEE Transactions on Signal Processing. – 2010. – Vol. 58, No. 11. – P. 5655-5666.
19. *Iftene S.* General secret sharing based on the chinese remainder theorem with applications in e-voting // Electronic Notes in Theoretical Computer Science. – 2007. – Vol. 186. – P. 67-84.
20. *Iliev A., Kyurkchiev N.* The faster extended Euclidean algorithm // Collection of scientific works from conference. – 2018. – P. 21-26.

REFERENCES

1. *Fan J., Vercauteren F.* Somewhat practical fully homomorphic encryption, *Cryptology ePrint Archive,* 2012.
2. *Gentry C., Sahai A., Waters B.* Homomorphic encryption from learning with errors: Conceptually-simpler, asymptotically-faster, attribute-based, *Advances in Cryptology–CRYPTO 2013: 33rd Annual Cryptology Conference, Santa Barbara, CA, USA, August 18-22, 2013. Proceedings, Part I*. Springer Berlin Heidelberg, 2013, pp. 75-92.
3. *Brakerski Z.* Fully homomorphic encryption without modulus switching from classical GapSVP, *Annual cryptology conference*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 868-886.
4. *Brakerski Z., Gentry C., Vaikuntanathan V.* (Leveled) fully homomorphic encryption without bootstrapping, *ACM Transactions on Computation Theory (TOCT),* 2014, Vol. 6, No. 3, pp. 1-36.

5. *Domingo-Ferrer J.* A provably secure additive and multiplicative privacy homomorphism, *International Conference on Information Security*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2002, pp. 471-483.
6. *Hariss K., Noura H., Samhat A. E.* An efficient fully homomorphic symmetric encryption algorithm, *Multimedia Tools and Applications,* 2020, Vol. 79, No. 17, pp. 12139-12164.
7. *Wang H., Wang Z., Domingo-Ferrer J.* Anonymous and secure aggregation scheme in fog-based public cloud computing, *Future Generation Computer Systems,* 2018, Vol. 78, pp. 712-719.
8. *Maqsood F. et al.* Cryptography: a comparative analysis for modern techniques, *International Journal of Advanced Computer Science and Applications,* 2017, Vol. 8, No. 6.
9. *Trepacheva A.V.* Uluchshennaya ataka po izvestnym otkrytym tekstam na gomomorfnuyu kriptosistemu Domingo-Ferrera [An Improved Known-Plaintext Attack on the Domingo-Ferrer Homomorphic Cryptosystem], *Tr. Instituta sistemnogo programmirovaniya RAN* [Proceedings of the Institute for System Programming of the Russian Academy of Sciences], 2014, Vol. 26, No. 5, pp. 83-98.
10. *Alabdulatif A., Kaosar M.* Privacy preserving cloud computation using Domingo-Ferrer scheme, *Journal of King Saud University-Computer and Information Sciences,* 2016, Vol. 28, No. 1, pp. 27-36.
11. *Merkel W. et al.* Factorization of numbers with physical systems, *Fortschritte der Physik: Progress of Physics,* 2006, Vol. 54, No. 8-10, pp. 856-865.
12. *Lenstra A. K. et al.* The factorization of the ninth Fermat number, *Mathematics of Computation,* 1993, Vol. 61, No. 203, pp. 319-349.
13. *Yakovlev V.A., Shemyakin S.N., Tarov E.V.* Ispol'zovanie metoda Montgomeri v algoritme bystrogo vozvedeniya v stepen' [Using Montgomery's method in a fast exponentiation algorithm], *I-methods,* 2023, Vol. 15, No. 1, pp. 6.
14. *Hossain M. A. et al.* Performance analysis of different cryptography algorithms, *International Journal of Advanced Research in Computer Science and Software Engineering,* 2016, Vol. 6, No. 3.
15. *Agner F.* Optimizing software in C++: An optimization guide for Windows, Linux and Mac platforms, 2020.
16. *Pei D., Salomaa A., Ding C.* Chinese remainder theorem: applications in computing, coding, cryptography. World Scientific, 1996.
17. *Schindler W.* A timing attack against RSA with the chinese remainder theorem, *Cryptographic Hardware and Embedded Systems—CHES 2000: Second International Workshop Worcester, MA, USA, August 17–18, 2000 Proceedings 2*. Springer Berlin Heidelberg, 2000, pp. 109-124.
18. *Wang W., Xia X. G.* A closed-form robust Chinese remainder theorem and its performance analysis, *IEEE Transactions on Signal Processing,* 2010, Vol. 58, No. 11, pp. 5655-5666.
19. *Iftene S.* General secret sharing based on the chinese remainder theorem with applications in e-voting, *Electronic Notes in Theoretical Computer Science,* 2007, Vol. 186, pp. 67-84.
20. *Iliev A., Kyurkchiev N.* The faster extended Euclidean algorithm, *Collection of scientific works from conference,* 2018, pp. 21-26.

Статью рекомендовал к опубликованию д.ф.-м.н., профессор В.О. Осипян.

**Бабенко Людмила Климентьевна** – Южный федеральный университет; e-mail: lkbabenko@sfedu.ru; тел.: +79054530191; г. Таганрог, Россия; кафедра безопасности информационных технологий им. Макаревича О.Б.; д.т.н.; профессор.

**Стародубцев Виталий Сергеевич** – e-mail: vstarodubcev@sfedu.ru; тел.: +79996928150; кафедра безопасности информационных технологий им. Макаревича О.Б.; аспирант.

**Babenko Lyudmila Kliment'evna** – Southern Federal University; e-mail: lkbabenko@sfedu.ru; phone: +79054530191; Taganrog, Russia; the Department of Information Technology Security named after Makarevich O.B.; dr of eng. sc.; professor.

**Starodubcev Vitalij Sergeevich** – e-mail: vstarodubcev@sfedu.ru; phone: +79996928150; the Department of Information Technology Security named after Makarevich O.B.; post graduate student.

УДК 519.711.3 DOI 10.18522/2311-3103-2024-5-16-28

**А.А. Кабанов, В.А. Крамарь, К.В. Дементьев**

## АЛГОРИТМ ОПТИМАЛЬНОГО КОМПЛЕКСИРОВАНИЯ ОЦЕНКИ СОСТОЯНИЙ В ДИСКРЕТНО-НЕПРЕРЫВНЫХ СИСТЕМАХ АНПА

*Статья посвящена разработке алгоритма оптимального комплексирования оценок состояний в дискретно-непрерывных системах. Целью исследования является создание эффективного метода объединения данных, получаемых от непрерывных и дискретных источников информации, для повышения точности и надежности оценки состояния сложных динамических систем. В статье подробно рассматриваются теоретические основы предложенного метода, включая математическое описание непрерывной и дискретной моделей системы, формулировку критерия оптимальности и вывод уравнений для вычисления весовых коэффициентов комплексирования. Особое внимание уделяется анализу условий, при которых предложенный алгоритм обеспечивает улучшение точности оценки по сравнению с использованием только непрерывного или только дискретного фильтра. Авторы приводят результаты численного моделирования, демонстрирующие эффективность разработанного алгоритма на примере оценки параметров движения автономного подводного аппарата. Показано, что предложенный метод комплексирования позволяет существенно снизить ошибки оценивания по сравнению с использованием отдельных фильтров, особенно в условиях неполноты и зашумленности измерений. В заключение делаются выводы о перспективности применения разработанного алгоритма в различных областях, связанных с обработкой информации в сложных технических системах, таких как навигация, управление движением, мониторинг состояния объектов и процессов. Отмечается, что предложенный подход может быть обобщен на случай комплексирования данных от большего числа источников информации и адаптирован к различным типам дискретно-непрерывных систем. Статья представляет интерес для специалистов в области теории управления, обработки сигналов и информации, а также разработчиков систем навигации и управления движением. Результаты исследования могут найти практическое применение при создании высокоточных систем оценивания состояния в различных технических приложениях.*

*Непрерывно-дискретные системы; комплексирование датчиков; оценка состояния; оптимальная оценка; терминальные системы.*

**A.A. Kabanov, V.A. Kramar, K.V. Dementiev**

## THE OPTIMAL STATE ESTIMATION FUSION ALGORITHM IN DISCRETE-CONTINUOUS AUV SYSTEMS

*The article is focused on the optimal state estimation fusion algorithm for discrete-continuous systems. The aim of the study is to create an effective fusion strategy for combining data obtained from continuous and discrete information sources to improve the accuracy and reliability of state estimation in complex dynamic systems. The paper discusses in detail the theoretical foundations of the proposed method, including the mathematical description of the continuous and discrete system models, the optimization criterion formulation, and the derivation of equations for calculating the complementation weights. Particular attention is paid to the analysis of conditions under which the proposed algorithm provides an improvement in estimation accuracy compared to the use of only continuous or only discrete filter. The authors present the results of numerical modeling demonstrating the developed algorithm efficiency on the example of autonomous underwater vehicle motion parameter estimation. It is shown that the proposed fusion method allows to significantly reduce the estimation errors compared to the use of separate filters, especially in conditions of incompleteness and noise in measurements. In conclusion, it is stated that the developed algorithm is promising for application in various fields related to information processing in complex technical systems, such as navigation, motion control, monitoring of the objects and processes. It is noted that the proposed approach can be generalized to the case of complexing data from a larger number of information sources and adapted to different types of discrete-continuous systems. The article is considered to be valuable for specialists in control theory, signal and information processing, as well as for developers of navigation and motion control systems. The research results can find practical application in the creation of high-precision state estimation systems in various technical applications.*

*Continuous-discrete systems; sensors fusion; state estimation; optimal estimation; terminal systems.*

**Введение.** Повышение требований точности к современным системам управления, функционирующих в условиях случайных возмущений, приводит к необходимости применения нескольких каналов измерений. В случае использования непрерывных и дискретных измерительных подсистем возникает проблемы их комплексирования с целью повышения качества и точности системы в целом. Также смешанные непрерывно-дискретные измерения характерны для современных средств автономной навигации.

Современные требования к движению автономных необитаемых подводных аппаратов (АНПА) характеризуется жесткими ограничениями на величины отклонения АНПА от заданной траектории. При этом существенно повышаются требования к точности определения координат АНПА [1, 2]. В этой связи все большее применение находят комплексные методы определения координат местоположения АНПА на базе объединения и оптимальной обработки навигационной информации, способствующих существенному повышению точности движения АНПА [3–6].

Одной из основных задач является оптимальное комплексирование данных от различных источников навигационной информации. При комплексном использовании навигационных средств для определения места АНПА должно быть обращено особое внимание на удельный вес каждого из источников информации. Оптимальное в смысле точности определения координат должно дать объединение измерений от всех источников, а не выделение, как в классическом случае, двух подсистем.

Как известно, одним из мощных методов решения такого рода задач является фильтрация Калмана [7–13]. Однако применение этого метода в данной задаче усложнено тем, что измерители функционируют как непрерывно, так и дискретно во времени [14].

Одним из классов нестационарных систем управления, в котором возникает задача комплексирования измерительных каналов непрерывного и дискретного типа, является класс систем управления конечного состояния [15]. Основной отличительной чертой таких систем управления является наличие терминальных условий (наличие жестких ограничений на значения фазовых координат системы в конечный момент времени).

Таким образом, в статье рассматривается задача управлением объектом, содержащим в контуре управления измерительные устройства, функционирующие как непрерывно, так и дискретно во времени. Необходимо объединить результаты таких измерений на основе калмановской фильтрации в единую систему с целью повышения точностных характеристик всего контура управления.

**Общая постановка задачи комплексирования.** Для систем управления, содержащих два канала измерений (непрерывный и дискретный), характерно одинаковое формализованное представление протекающих в контуре физических процессов.

Модель динамики систем представляется взаимосвязанной системой обыкновенных дифференциальных и конечно-разностных уравнений вида

$$\begin{gathered}\frac{dx(t)}{dt} = A_1(t)x(t) + B_1(t)u_1(t) + \omega_1(t), \quad t \in T_i, \\ x(t_{i+1}) = A_2(t_i)x(t_i) + B_2(t_i)u_2(t_i) + \omega_2(t_i), \quad t_i \in \Theta, \\ x(t_0) = x_0.\end{gathered} \tag{1}$$

В уравнении (1) приняты следующие обозначения: $T_i = [t_i, t_{i+1})$ – полуоткрытый временной интервал функционирования системы управления между моментами работы дискретной подсистемы, причем $T = \left[t_0, t_f\right]$ – полное время функционирования системы, а $t_0$ и $t_f$ – соответственно начальный и конечный моменты времени процесса управления, $\Theta = \left\{t_1, t_2, \dots, t_{k_f} : (t_{i+1} - t_i) > 0\right\}$ – множество моментов функционирования дискретной подсистемы, $x \in R^n$ – вектор состояния системы, элементы которого является непрерывно-дифференцируемым на $T_i$ функциями и терпят разрывы первого рода на $\Theta$, причем за значения элементов вектора $x$ в точках разрыва принимаются правые пределы, $u_1 \in R^{m_1}$, $u_2 \in R^{m_2}$ – векторы управления непрерывной и дискретной подсистем соответственно, $\omega_1$ и $\omega_2$ – $n$-мерные векторы возмущений, $x_0$ – вектор начального состояния системы.

Система управления содержит непрерывные и дискретные измерительные устройства, функционирующие в соответствие с уравнениями

$$\begin{aligned} y_1(t) &= C_1(t)x(t) + \nu_1(t), \quad t \in T_i, \\ y_2(t_i) &= C_2(t_i)x(t_i) + \nu_2(t_i), \quad t_i \in \Theta, \end{aligned} \tag{2}$$

где $y_1(t)$ и $y_2(t_i)$ – $q_1$- и $q_2$-мерные векторы непрерывных и дискретных измерений, $\nu_1(t)$ и $\nu_2(t_i)$ – векторы ошибок измерений соответствующей размерности.

Матрицы $A_1$, $B_1$, $C_1$ и $A_2$, $B_2$, $C_2$ соответственно состоят из кусочно-непрерывных и кусочно-постоянных элементов и согласованы с описанием (1), (2).

Задача комплексирования в общем виде формулируется следующим образом. Необходимо найти такое оптимальное объединение сигналов от непрерывных и дискретных измерителей, которое позволяет существенно повысить точностные характеристики всей системы управления.

В связи с тем, что реальные системы управления подвержены влиянию различного рода случайных факторов, действующих как на сам объект управления, так и на информационно измерительный комплекс, поставленную задачу можно сформулировать следующим образом. Требуется построить алгоритм оптимальной статистической обработки данных, позволяющий осуществить рациональное комплексирование отдельных измерительных каналов, с целью повышения точности системы управления на основе более точного определения текущих значений, измеряемых параметров.

Под оптимальной статистической обработкой будем понимать задачу синтеза оптимального оценщика вектора состояния системы на основе метода оптимального нестационарного оценивания состояния Калмана за счет их удобной реализации и возможности выработки текущей оценки фазовых координат системы управления в моменты поступления сигналов измерения.

Решение общей задачи требует применение обобщенного метода Калмана для непрерывно-дискретных систем с математическими моделями (1), (2). При этом особую роль играет характеристическое тождество Грина-Лагранжа, позволяющее эффективно решать задачи анализа и синтеза.

**Характеристическое тождество Грина–Лагранжа.** Будем рассматривать математическое описание непрерывно-дискретных систем управления (1). Запишем (1) в виде операторных уравнений

$$\begin{aligned} \mathcal{A}x(t) &= f_1(t), \quad t \in T_i, \\ \mathcal{A}x(t_i) &= f_2(t_i), \quad t_i \in \Theta, \\ x(t_0) &= x_0. \end{aligned} \tag{3}$$

где $\mathcal{A}x(t)$ и $\mathcal{A}x(t_i)$ – соответственно обыкновенные дифференциальный и конечноразностный линейные операторы $\mathcal{A}: T \to \mathbb{R}^n$ вида

$$\begin{aligned} \mathcal{A}x(t) &= \frac{dx(t)}{dt} - A_1(t)x(t), \\ \mathcal{A}x(t_i) &= x(t_{i+1}) - A_2(t_i)x(t_i), \end{aligned} \tag{4}$$

а $f_1$ и $f_2$ – $n$-векторы внешних воздействий, определенные как

$$\begin{aligned} f_1(t) &= B_1(t)u_1(t) + w_1(t), \\ f_2(t_i) &= B_2(t_i)u_2(t_i) + w_2(t_i). \end{aligned} \tag{5}$$

Такое представление математической модели непрерывно-дискретной системы можно трактовать как уравнение движения системы, описываемое первым уравнением системы (3), с многоточечными краевыми условиями, следующими из второго уравнения той же системы. Другими словами, конечно-разностные уравнения системы (3) определяют величину разрыва траектории системы.

По своему математическому описанию системы вида (3) принадлежат классу функционально-сложных динамических систем гриновского типа [16, 17]. Основой таких систем является применение общей методологии Грина - Лагранжа для формирования сопряженных операторов и билинейных функционалов как обобщенных элементов характеризации динамики исследуемых систем. В качестве аппарата решения задач анализа и синтеза рассматриваемых систем используется формализм Грина-Лагранжа, в основе которого лежит характеристическое тождество системы или формула Грина

$$\int_{t_1}^{t_2}[X^T(t)(\mathcal{A}Y)(t)-(\mathcal{A}^*X^T)(t)Y(t)]dt = L_{\mathcal{A}}[X^T,Y](t)\Big|_{t_1}^{t_2}.$$

Конструирование этого тождества порождает по линейному оператору системы $\mathcal{A}$ и согласованными с ним матричными функциями $Y(t)$ и $X(t)$ два новых объекта: линейный оператор $\mathcal{A}^*$, сопряженный в смысле Лагранжа оператору $\mathcal{A}$, и билинейный матричный функционал $L_{\mathcal{A}}$ (конкомитант Лагранжа). Если для линейного оператора $\mathcal{A}$ можно построить такое тождество, то оно называется гриновским. Значение билинейного матричного функционала $L_{\mathcal{A}}$ в момент времени $t$ определяется значениями матричных функций $Y(\cdot)$ слева и $X(\cdot)$ справа от $t$ соответственно, что является следствием свойства неупреждаемости (причинности) оператора $\mathcal{A}$, характерного для моделей реальных динамических систем и процессов. Прямой $\mathcal{A}$ и сопряженный $\mathcal{A}^*$ операторы, билинейный функционал $L_{\mathcal{A}}$, их динамические свойства полностью характеризуют исследуемую систему и в этом смысле называются элементами характеризации. Связаны эти элементы описания системы характеристическим тождеством. Оно является конструктивным объектом, т.е. с его помощью можно строить элементы характеризации для конкретных типов линейных функционально-сложных динамических систем.

Тождество Грина является характеристическим в том смысле, что оно по заданному оператору системы $\mathcal{A}$ определяет сопряженный оператор и билинейный функционал, являющиеся обобщенными элементами характеризации системы.

Построим характеристическое тождество для непрерывно-дискретных систем, заданных в форме (3), (4), используя правило множителей Лагранжа [18]. Для этого умножим операторы $(\mathcal{A}x)(\tau)$ и $(\mathcal{A}x)(t_i)$ слева соответственно на матрицы $Z(\tau)$ и $Z(t_i)$ размерности $n\times n$, вид которых определяется ниже. Первое произведение проинтегрируем по $\tau$ в пределах от $t_0$ до $t$, а второе просуммируем по $i$ от 1 до $k(t)$, где $k(t)=\max\{m: t_m\le t, t_m\in\Theta\}$. Сложим полученные результаты и образуем функционал

$$\mathcal{J}=\int_{t_0}^{t}Z(\tau)(\mathcal{A}x)(\tau)d\tau+\sum_{i=1}^{k(t)}Z(t_i)(\mathcal{A}x)(t_i). \tag{6}$$

Осуществим ряд тождественных преобразований, введя обозначение для сопряженного с $(\mathcal{A}x)(\tau)$ оператора $(\mathcal{A}^*Z)(\tau)$ и запишем функционал (6) в виде

$$\mathcal{J}=\int_{t_0}^{t}(\mathcal{A}^*Z)(\tau)x(\tau)d\tau+\sum_{i=1}^{k(t)}(\mathcal{A}^*Z)(t_{i-1})x(t_{i-1})+Z(\tau)x(\tau)|_{t_0}^{t}, \tag{7}$$

где

$$(\mathcal{A}^*Z)(\tau)=-\frac{d}{d\tau}Z(\tau)-Z(\tau)A_1(\tau), \tag{8}$$

$$(\mathcal{A}^*Z)(t_{i-1})=Z(t_{i-1})-Z(t_i)A_2(t_i). \tag{9}$$

Запишем характеристическое тождество Грина-Лагранжа для непрерывно-дискретных систем, математическая модель которых задана уравнениями (3), (4), (5)

$$\begin{gathered}\int_{t_0}^{t} Z(\tau)(\mathcal{A}x)(\tau)d\tau + \sum_{i=1}^{k(t)} Z(t_i)(\mathcal{A}x)(t_i) - \\ \int_{t_0}^{t} (\mathcal{A}^* Z)(\tau)x(\tau)d\tau + \sum_{i=1}^{k(t)} (\mathcal{A}^* Z)(t_{i-1})x(t_{i-1}) = \\ = \beta(Z,x)(t) - \beta(Z,x)(t_0).\end{gathered} \tag{10}$$

В соотношении (9) $\beta(Z,x)(\cdot) = Z(\cdot)x(\cdot)$ – билинейная форма.

Тождество (10) является характеристическим тождеством Грина-Лагранжа для непрерывно-дискретных систем, математическая модель которых задана операторными уравнениями (3) с определением исходных параметров (4).

Матрица $Z$ определяется однородными уравнениями на полуинтервалах $T_i$ и на последовательности с помощью сопряженного оператора $\mathcal{A}^*: T \to \mathbb{R}^{n\times n}$. Также требуется, чтобы матрица $Z$ в момент времени $t$ обращалась в единичную. В результате получаем описание

$$\begin{gathered}(\mathcal{A}^* Z)(\tau) = 0,\ \tau \in T_i, \\ (\mathcal{A}^* Z)(t_{i-1}) = 0,\ t_i \in \Theta, \\ Z(t) = I_n.\end{gathered} \tag{11}$$

Поскольку матрица $Z$, удовлетворяющая уравнениям (11) при $\tau = t$ принимает фиксированное значение, то она по существу является функцией двух переменных – текущего аргумента $\tau$, $t_0 \le \tau < t$ и момента наблюдения за поведением системы $t$. Желая это подчеркнуть, для матрицы $Z$ можно ввести обозначение $G(t,\tau)$.

С учетом свойств матрицы $Z$ из характеристического тождества (10) следует представление решений непрерывно-дискретной системы (3) в виде аналога формулы Коши из теории обыкновенных дифференциальных уравнений

$$x(t) = G(t,t_0)x(t_0) + \int_{t_0}^{t} G(t,\tau)f_1(\tau)d\tau + \sum_{i=1}^{k(t)} G(t,t_i)f_2(t_i), \tag{12}$$

где $G(t,\cdot)$ – матрица весовых функций системы или матричная функция Грина, причем $G(t,\cdot) \in \mathbb{R}^{n\times n}$.

**Оптимальное оценивание состояния линейных нестационарных непрерывно-дискретных систем.** Задача оптимального оценивания состояния линейных нестационарных непрерывно-дискретных систем математически эквивалентны детерминированной задаче оптимального управления сопряженной системой по квадратическому критерию качества.

Сформулируем задачу фильтрации, являющуюся обобщением на непрерывно-дискретные системы задачи Калмана.

Пусть рассматривается линейный нестационарный непрерывно-дискретный объект, заданный операторными уравнениями

$$\begin{gathered}\mathcal{A}x(t) = B_1(t)w_1(t) + E_1(t)\mathbb{f}_1(t), \quad t \in T_i, \\ \mathcal{A}x(t_i) = B_2(t_i)w_2(t_i) + E_2(t_i)\mathbb{f}_2(t_i), \quad t_i \in \Theta, \\ x(t_0) = x_0.\end{gathered} \tag{13}$$

где операторы $\mathcal{A}: T \to \mathbb{R}^n$ имеют вид (4).

В соотношении (13) $w_1$ и $w_2$ – непрерывный и дискретный белые шумы, а $\mathbb{f}_1$ и $\mathbb{f}_2$ – непрерывные и дискретные управляющие, либо возмущающие детерминированные воздействия соответственно.

Пусть также известны уравнения наблюдений за поведением системы

$$\begin{gathered}y_1(t) = C_1(t)x(t) + v_1(t) + \xi_1(t), \quad t \in T_i, \\ y_2(t_i) = C_2(t_i)x(t_i) + v_2(t_i) + \xi_2(t_i), \quad t_i \in \Theta,\end{gathered} \tag{14}$$

где $\vartheta_1$ и $\vartheta_2$ – соответственно непрерывный и дискретный белые шумы, характеризующие случайные флуктуации измерений, а $\xi_1$ и $\xi_2$ – известные детерминированные непрерывная и дискретная величины, появление которых в уравнениях измерений обусловлено систематическими погрешностями измерений.

Предполагаем, что $x_0$ гауссовый случайный вектор с характеристиками

$$\begin{gathered} E[x(t_0)] = m_x(t_0), \\ E[(x(t_0) - m_x(t_0))(x(t_0) - m_x(t_0))^T] = P_0, \end{gathered} \tag{15}$$

где $E$ – оператор математического ожидания. Также известны статистические характеристики гауссовых случайных процессов и последовательностей, входящих в уравнение (13), (14)

$$E[w_1(t)] = m_{w_1}(t), E\left[\left(w_1(t) - m_{w_1}(t)\right)\left(w_1(\tau) - m_{w_1}(\tau)\right)^T\right] = Q_1(t)\delta(t-\tau),$$
$$E[w_2(t_i)] = m_{w_2}(t_i), E\left[\left(w_2(t_i) - m_{w_2}(t_i)\right)\left(w_2(t_j) - m_{w_2}(t_j)\right)^T\right] = Q_2(t_j)\delta_{ij},$$
$$E[\vartheta_1(t)] = m_{\vartheta_1}(t), E\left[\left(\vartheta_1(t) - m_{\vartheta_1}(t)\right)\left(\vartheta_1(\tau) - m_{\vartheta_1}(\tau)\right)^T\right] = R_1(t)\delta(t-\tau),$$
$$E[\vartheta_2(t_i)] = m_{\vartheta_2}(t_i), E\left[\left(\vartheta_2(t_i) - m_{\vartheta_2}(t_i)\right)\left(\vartheta_2(t_j) - m_{\vartheta_2}(t_j)\right)^T\right] = R_2(t_j)\delta_{ij},$$

$$E\left[\left(w_1(t) - m_{w_1}(t)\right)\left(\vartheta_1(\tau) - m_{\vartheta_1}(\tau)\right)^T\right] = S_1(t)\delta(t-\tau),$$
$$E\left[\left(w_2(t_i) - m_{w_2}(t_i)\right)\left(\vartheta_2(t_j) - m_{\vartheta_2}(t_j)\right)^T\right] = S_2(t_j)\delta_{ij},$$
$$E\left[\left(w_1(t) - m_{w_1}(t)\right)\left(w_2(t_i) - m_{w_2}(t_i)\right)^T\right] = 0,$$
$$E\left[\left(w_1(t) - m_{w_1}(t)\right)\left(\vartheta_2(t_i) - m_{\vartheta_2}(t_i)\right)^T\right] = 0,$$
$$E\left[\left(\vartheta_1(t) - m_{\vartheta_1}(t)\right)\left(w_2(t_i) - m_{w_2}(t_i)\right)^T\right] = 0,$$
$$E\left[\left(\vartheta_1(t_i) - m_{\vartheta_1}(t_i)\right)\left(\vartheta_2(t_j) - m_{\vartheta_2}(t_j)\right)^T\right] = 0,$$

где $\delta_{ij}$ – символ Кронекера, матрицы $P_0$, $Q_1$, $Q_2$ симметричны и положительно полуопределены, а матрицы $R_1$, $R_2$ симметричны и положительно определены. Начальное состояние $x_0$ некоррелировано с шумами $\vartheta_1$, $\vartheta_2$, $w_1$ и $w_2$.

Задачу фильтрации по аналогии с [19] сформулируем следующим образом. Необходимо по результатам измерений (14) построить оптимальную оценку $\hat{x}$ вектора состояния системы в любой произвольный фиксированный момент времени $\Theta \geq t_0$.

Оценку будем искать в классе линейных оценок вида

$$a^T\hat{x}(\Theta) = b^T m_x(t_0) - \int_{t_0} u_1^T(t)y_1(t)dt - \sum_{i=1}^{k(\Theta)} u_2^T(t_i)y_2(t_i), \tag{16}$$

обеспечивающих минимум функционалу

$$\mathcal{J} = E[\{a^T[x(\theta) - \hat{x}(\ )]\}^2]. \tag{17}$$

Выражение (16) определяет структуру непрерывно-дискретной системы, на вход которой поступают сигналы от измерительных устройств $y_1, y_2$, а на выходе формируется оценка вектора состояний системы (13).

Задачу также можно сформулировать следующим образом. Необходимо определить, каким условиям подчиняются векторы $b, u_1, u_2$, обеспечивающие оптимальность оценке $\hat{x}$.

Используя уравнения измерений (14) и второе уравнение системы (13) преобразуем выражение для оценки (16) (далее аргументы для матриц системы (13) и уравнений (15) при необходимости, для простоты записи будем опускать)

$$a^T\hat{x}(\Theta) = b^T m_x(t_0) - \int_{t_0}^{\Theta} u_1^T(t)[C_1x(t) + \nu_1(t) + \xi_1(t)]\,dt - \\ - \sum_{i=1}^{k(\Theta)} u_2^T(t_i)[C_2A_2x(t_{i-1}) + C_2B_2w_2(t_i) + C_2E_2\mathbb{f}_2(t_{i-1}) + \nu_2(t_i) + \xi_2(t_i)]\,. \tag{18}$$

Представляя характеристическое тождество Грина-Лагранжа для системы (13) в виде

$$z^T(\Theta)x(\Theta) = z^T(t_0)x(t_0) + \int_{t_0}^{\Theta}[z^T(t)B_1w_1(t) + z^T(t)E_1\mathbb{f}_1(t) - (\mathcal{A}^*z^T)x(t)]dt + \\ + \sum_{i=1}^{k(\Theta)}[z^T(t_i)B_2w_2(t_i) + z^T(t_i)E_2\mathbb{f}_2(t_{i-1}) - (\mathcal{A}^*z^T)(t_{i-1})x(t_{i-1})], \tag{19}$$

и вычитая из него соотношение (18) получаем

$$z^T(\ )x(\ ) - a^T\hat{x}(\ ) = z^T(t_0)x(t_0) - b^Tm_x(t_0) + \int_{t_0}[-(\mathcal{A}^*z^T)x(t) + u_1^T(t)C_1]x(t)dt + \\ + \sum_{i=1}^{k(\ )}[-(\mathcal{A}^*z^T)(t_{i-1})x(t_{i-1}) + u_2^T(t_i)C_2\mathcal{A}_2]x(t_{i-1}) + \\ + \int_{t_0}[z^T(t)B_1w_1(t) + z^T(t)E_1\mathbb{f}_1(t) + u_1^T(t)\nu_1(t) + u_1^T(t)\xi_1(t)]dt \\ + \sum_{i=1}^{k(\ )}[z^T(t_i)B_2w_2(t_i) + z^T(t_i)E_2\mathbb{f}_2(t_{i-1}) + u_2^T(t_i)C_2B_2(t_i) + \\ + u_2^T(t_i)C_2E_2\mathbb{f}_2(t_{i-1})\nu_2(t_i) + u_2^T(t_i)\xi_2(t_i)]. \tag{20}$$

Определим сопряженную непрерывно-дискретную систему следующим образом

$$(\mathcal{A}^*z^T)(t) = u_1^T(t)C_1(t), \quad t \in T_i, \\ (\mathcal{A}^*z^T)(t_{i-1}) = u_2^T(t_i)C_2(t_i)A_2(t_i), \quad t_i \in \Theta, \\ z^T(\ ) = a^T. \tag{21}$$

Тогда выражение (20) примет вид

$$a^T[x(\theta) - \hat{x}(\theta)] = z^T(t_0)x(t_0) - b^Tm_x(t_0) + \\ + \int_{t_0}[z^T(t)B_1w_1(t) + z^T(t)E_1\mathbb{f}_1(t) + u_1^T(t)\nu_1(t) + u_1^T(t)\xi_1(t)]dt + \\ + \sum_{i=1}^{k(\ )}[z^T(t_i)B_2w_2(t_i) + z^T(t_i)E_2\mathbb{f}_2(t_{i-1}) + u_2^T(t_i)C_2B_2(t_i) + \\ + u_2^T(t_i)C_2E_2\mathbb{f}_2(t_{i-1}) + u_2^T(t_i)\nu_2(t_i) + u_2^T(t_i)\xi_2(t_i)]. \tag{22}$$

Применяя к обеим частям (22) операцию математического ожидания, можно определить условие безусловной несмещенности оценки $\hat{x}$

$$E[a^T\{x(\ ) - \hat{x}(\ )\}] = 0$$

в форме

$$b^Tm_x(t_0) = z^T(t_0)m_x(t_0) + \int_{t_0}\{z^T(t)\big[B_1m_{w_1}(t) + E_1\mathbb{f}_1(t)\big] + u_1^T(t)\big[m_{\nu_1}(t) + \xi_1(t)\big]\}dt + \\ + \sum_{i=1}^{k(\ )}\{z^T(t_i)\big[B_2m_{w_2}(t_i) + E_2\mathbb{f}_2(t_{i-1})\big] + \\ + u_2^T(t_i)\big[C_2B_2m_{w_2}(t_i) + C_2E_2\mathbb{f}_2(t_{i-1}) + m_{\nu_2}(t_i) + \xi_2(t_i)\big]\}, \tag{23}$$

что фактически накладывает условие на вектор $b$.

Учитывая (22) и (23) и выполняя преобразования с учетом априорных данных сформируем функционал в виде

$$\mathcal{J} = z^T(t_0)P_0z(t_0) + \int_{t_0}[z^T(t)\mathcal{Q}_1(t)z(t) + 2z^T(t)\mathcal{Q}_{11}^T(t)u_1(t) + u_1^T(t)\mathcal{R}_1(t)u_1(t)]dt + \\ + \sum_{i=1}^{k(\ )}[z^T(t_i)\mathcal{Q}_2(t_i)z(t_i) + 2z^T(t_i)\mathcal{Q}_{22}^T(t_i)u_2(t_i) + u_2^T(t_i)\mathcal{R}_2(t_i)u_2(t_i)], \tag{24}$$

где

$$\begin{gathered} Q_1(t) = B_1(t)P_{11}(t)B_1^T(t), \quad Q_{11}(t) = S_1^T(t)B_1^T(t), \quad \mathcal{R}_1(t) = P_{21}(t), \\ Q_2(t_i) = B_2(t_i)P_{12}(t_i)B_2^T(t_i), \quad Q_{22}(t_i) = S^T(t_i)B_2^T(t_i)P_{12}(t_i)B_2^T(t_i), \\ \mathcal{R}_2(t_i) = C_2(t_i)B_2(t_i)P_{12}(t_i)C_2^T(t_i) + C_2(t_i)B_2(t_i)S_2(t_i) + \\ +S_2^T(t_i)B_2^T(t_i)C_2^T(t_i) + P_{22}(t_i). \end{gathered} \tag{25}$$

Предположим, что все взаимные и автоковариационные функции, входящие в выражении для матрицы $\mathcal{R}_2$ таковы, что эта матрица является несингулярной.

Для того, чтобы задача (21), (24) принадлежала к классу линейно-квадратических задач управления, должны выполняться условия для определяющих матриц [20]

$$P_0 \geq 0, \begin{bmatrix} Q_1(t) & Q_{11}^T(t) \\ Q_{11}(t) & \mathcal{R}_1(t) \end{bmatrix} \geq 0, \begin{bmatrix} Q_2(t_i) & Q_{22}^T(t_i) \\ Q_{22}(t_i) & \mathcal{R}_2(t_i) \end{bmatrix} \geq 0. \tag{26}$$

Указанные условия выполняются. Первое условие выполняется, поскольку по постановке задачи матрица $P_0$ положительно полуопределена. Второе и третье условие доказывается, основываясь на том, что ковариационная матрица гауссова вектора положительно полуопределена.

Таким образом, приходим к задаче построения оптимального управления $u_1, u_2$ сопряженной с (13) непрерывно-дискретной линейной нестационарной детерминированной системой (21), доставляющий абсолютный минимум квадратическому функционалу (24). Это отражает принцип двойственности между задачами оптимальной линейной фильтрации и детерминированного линейно-квадратического оптимального управления в непрерывно-дискретных нестационарных системах.

Далее перейдем непосредственно к построению оптимального линейного нестационарного непрерывно-дискретного фильтра.

Прибавляя к функционалу (24) вспомогательное тождество метода дополнения до полного квадрата, развитого Лежандром применительно к задаче изучения вторых вариаций в вариационном управлении [21],

$$\begin{gathered} \int_{t_0}^{t_f} \frac{d}{dt}[x^T(t)K(t)x(t)]dt + \sum_{i=1}^{k_f}[x^T(t_i)K(t_i)x(t_i) - x^T(t_{i-1})K(t_{i-1})x(t_{i-1})] = \\ = x^T(t_f)K(t_f)x(t_f) - x^T(t_0)K(t_0)x(t_0), \end{gathered}$$

в которое вместо $x(t)$, $K(t)$, $t_f$, $K_f$ подставляем соответственно $z(t)$, $P(t)$, $\theta$, $k(\theta)$, определяя матрицу ковариации ошибки оценивания $P(t)$ на решениях системы

$$\begin{gathered} \frac{d}{dt}P(t) = -A_1(t)P(t) - P(t)A_1^T(t) + Q_1(t) - \tilde{Q}_{11}^T(t)\mathcal{R}_1^{-1}(t)\tilde{Q}_{11}^T(t), \ t \in T_i, \\ P(t_i) = A_2(t_i)P(t_{i-1})A_2^T(t_i) + Q_2(t_i) - \tilde{Q}_{22}(t_i)\tilde{\mathcal{R}}_2^{-1}(t_i)\tilde{Q}_{22}(t_i), \quad t_i \in \Theta, \\ P(t_0) = P_0, \end{gathered} \tag{27}$$

выполняя преобразования, получим функционал в виде

$$\begin{gathered} J = z^T(\ )P(\ )z(\ ) + \int_{t_0}\left[u_1(t) + \mathcal{R}_1^{-1}(t)\tilde{Q}_{11}z(t)\right]^T\mathcal{R}_1(t)\left[u_1(t) + \mathcal{R}_1^{-1}(t)\tilde{Q}_{11}(t)z(t)\right]dt + \\ + \sum_{i=1}^{k(\ )}\left[u_2(t_i) + \tilde{\mathcal{R}}_2^{-1}(t_i)\tilde{Q}_{22}(t_i)z(t_i)\right]^T\tilde{\mathcal{R}}_2(t_i)\left[u_2(t_i) + \tilde{\mathcal{R}}_2^{-1}(t_i)\tilde{Q}_{22}(t_i)z(t_i)\right], \end{gathered} \tag{28}$$

где

$$\begin{gathered} \tilde{Q}_{11}(t) = C_1(t)P(t) + Q_{11}(t), \tilde{Q}_{22}(t_i) = C_2(t_i)A_2(t_i)P(t_{i-1})A_2^T(t_i) + Q_{22}(t_i), \\ \tilde{\mathcal{R}}_2(t_i) = \mathcal{R}_2(t_i)C_2(t_i)A_2(t_i)P(t_{i-1})A_2^T(t_i)C_2^T(t_i) \end{gathered}$$

и определим оптимальное управление непрерывно-дискретной системой (21) в виде

$$\begin{gathered} u_1(t) - \mathcal{R}_1^{-1}(t)\tilde{Q}_{11}(t)z(t), t \in T_i, \\ u_2(t_i) = -\tilde{\mathcal{R}}_2^{-1}(t_i)\tilde{Q}_{22}(t_i)z(t_i), t_i \in \Theta. \end{gathered} \tag{29}$$

Минимальное значение функционала (24) определяется выражением

$$\mathcal{J}_{opt} = z^T(\ )P(\ )z(\ )\#(30)$$

Подставим полученное уравнение оптимального управления в уравнение системы (21)

$$\begin{gathered}(\mathcal{A}^* z^T)(t) = -z^T(t)\tilde{\mathcal{Q}}_{11}^T(t)\mathcal{R}_1^{-1}(t)C_1(t), t \in T_i,\\ (\mathcal{A}^* z^T)(t_{i-1}) = -z^T(t_i)\tilde{\mathcal{Q}}_{22}^T(t_i)\tilde{\mathcal{R}}_2^{-1}(t_i)C_2(t_i)A_2(t_i), t_i \in\ ,\\ z^T(\ ) = a^T\end{gathered} \quad (31)$$

и в представление оценки вектора состояния системы (13) с учетом условия безусловной несмещенности запишем характеристическое тождество Грина–Лагранжа для системы (28)

$$\begin{gathered}\int_{t_0}[z^T(t)(\mathcal{A}\hat{x})(t) - (\mathcal{A}^* z^T)(t)\hat{x}(t)]dt +\\ +\sum_{i=1}^{k(\ )}[z^T(t_i)(\mathcal{A}\hat{x})(t_i) - (\mathcal{A}^* z^T)(t_{i-1})\hat{x}(t_{i-1})] =\\ = z^T(\ )\hat{x}(\ ) - z^T(t_0)\hat{x}(t_0).\end{gathered} \quad (32)$$

Подставляя в него уравнение (31) в результате получаем

$$\begin{gathered}z^T(t_0)[\hat{x}(t_0) - m_x(t_0)] + \int_{t_0} z^T(t)\{(\mathcal{A}\hat{x})(t) - B_1 m_{w_1}(t) - E_1\mathbb{f}_1(t) -\\ -K(t)\big[y_1(t) - m_{v_1}(t) - \xi_1(t) - C_1\hat{x}(t)\big]\}dt +\\ +\sum_{i=1}^{k(\ )} z^T(t_i)\{(\mathcal{A}\hat{x})(t_i) - B_2 m_{w_2}(t_i) - E_2\mathbb{f}_2(t_{i-1}) - K(t_i)\times\\ \times\big[y_2(t_i) - C_2B_2 m_{w_2}(t_i) - C_2E_2\mathbb{f}_2(t_{i-1}) - m_{v_2}(t_i) - \xi_2(t_i) - C_2A_2\hat{x}(t_{i-1})\big]\} = 0,\end{gathered} \quad (33)$$

где

$$\begin{gathered}K(t) = [P(t)C_1^T(t) + B_1(t)S_1(t)]R_1^{-1}(t),\\ K(t_i) = [P_1(t_i)C_2^T(t_i) + B_2(t_i)S_2(t_i)][C_2(t_i)P_1(t_i)C_2^T(t_i) +\\ +C_2(t_i)B_2(t_i)S_2(t_i) + S_2^T B_2^T(t_i)C_2^T(t_i) + R_2(t_i)]^{-1},\end{gathered} \quad (34)$$

$$P_1(t_i) = A_2(t_i)P(t_{i-1})A_2^T(t_i) + B_2(t_i)Q_2(t_i)B_2^T(t_i). \quad (35)$$

Решение однородной непрерывно-дискретной системы (31) может быть представлено в виде

$$z(t) = G(t,\ )a, \qquad \forall t \in T, \quad (36)$$

где $G(t,\ )$ – матрица весовых функций системы (31).

Запишем представление оценки вектора состояния системы (13) с учетом условия безусловной несмещенности $\hat{x}(\ )$ в виде

$$\begin{gathered}a^T\hat{x}(\ ) = z^T(t_0)m_x(t_0) + \int_{t_0} z^T(t)[\Lambda_1(t) - \Pi_1(t)y_1(t)]dt +\\ +\sum_{i=1}^{k(\ )} z^T(t_i)[\Lambda_2(t_i) - \Pi_2(t_i)y_2(t_i)],\end{gathered} \quad (37)$$

где

$$\begin{gathered}\Lambda_1(t) = B_1 m_{w_1}(t) + E_1\mathbb{f}_1(t) - \Pi_1(t)\big[m_{v_1}(t) + \xi_1(t)\big],\\ \Pi_1(t) = \tilde{\mathcal{Q}}_{11}^T(t)\mathcal{R}_1^{-1}(t),\ \Pi_2(t_i) = \tilde{\mathcal{Q}}_{22}^T(t_i)\tilde{\mathcal{R}}_2^{-1}(t_i).\\ \Lambda_2(t_i) = B_2 m_{w_2}(t_i) + E_2\mathbb{f}_2(t_{i-1}) -\\ -\Pi_2(t_i)\big[C_2B_2 m_{w_2}(t_i) + C_2E_2\mathbb{f}_2(t_{i-1}) + m_{v_2}(t_i) + \xi_2(t_i)\big],\end{gathered} \quad (38)$$

Анализ (34) позволяет сделать вывод, что оптимальная оценка вектора состояния непрерывно-дискретной системы (13) $\hat{x}(\theta)$ единственна и не зависит от выбора вектора $a$. Тогда необходимым и достаточным условием выполнения (33) является уравнения

$$\begin{gathered}(A\hat{x})(t) = B_1(t)m_{w_1}(t) + E_1(t)\mathbb{f}_1(t) + \\ +K(t)\big[y_1(t) - m_{v_1}(t) - \xi_1(t) - C_1(t)\hat{x}(t)\big], t \in T_i, \\ (A\hat{x})(t_i) = B_2(t_i)m_{w_2}(t_i) - E_2(t_i)\mathbb{f}_2(t_{i-1}) + \\ +K(t_i)[y_2(t_i) - C_2(t_i)B_2(t_i)m_{w_2}(t_i) - C_2(t_i)E_2(t_i)\mathbb{f}_2(t_{i-1}) - \\ -m_{v_2}(t_i) - \xi_2(t_i) - C_2(t_i)A_2(t_i)\hat{x}(t_{i-1})], t_i \in \Theta, \\ \hat{x}(t_0) = m_x(t_0),\end{gathered} \tag{39}$$

которые определяют конструкцию оптимального линейного нестационарного непрерывно-дискретного фильтра.

Таким образом получаем следующий результат.

Пусть необходимо построить оптимальную оценку $\hat{x}$ вектора состояния $x$непрерывно-дискретной системы (13) по результатам наблюдений (14) с известными априорными характеристиками случайных гауссовых вектора начального состояния системы, процессов и последовательностей, входящих в уравнения (13), (14). При этом оценка $\hat{x}$ должна принадлежать классу линейных оценок вида (16) и обеспечивать абсолютный минимум функционалу (17). Тогда указанная оценка удовлетворяет системе (39), где матрица коэффициентов усиления фильтра $K$ и матрица ковариаций ошибки оценивания $P$ соответственно удовлетворяют уравнениям (34), (35), (36).

В том случае, когда все входящие в уравнения (13), (14) шумы центрированы, некоррелированы между собой и вектором начального состояния, а также отсутствует детерминированные слагаемые в правой части уравнений (13), (14), полученные уравнения упрощаются.

Тогда уравнения оптимального нестационарного непрерывно-дискретного фильтра будет иметь вид

$$\begin{gathered}\mathcal{A}\hat{x}(t) = K(t)[y_1(t) - C_1(t)\hat{x}(t)],\ t \in T_i, \\ \mathcal{A}\hat{x}(t_i) = K(t_i)[y_2(t_i) - C_2(t_i)A_2(t_i)\hat{x}(t_{i-1})],\ t_i \in \Theta, \\ x(t_0) = 0.\end{gathered} \tag{40}$$

Уравнения для матричных коэффициентов усиления фильтра записываются в виде

$$\begin{gathered}K(t) = P(t)C_1^T(t)R_1^{-1}(t), \quad t \in T_i, \\ K(t_i) = P_1(t_i)C_2^T(t_i)[C_2(t_i)P_1(t_i) + R_2(t_i)]^{-1}, \quad t_i \in \Theta,\end{gathered} \tag{41}$$

а уравнение для ковариационной матрицы ошибки оценивания имеют вид

$$\begin{gathered}\frac{dP(t)}{dt} = A_1(t)P(t) + P(t)A_1^T(t) + B_1(t)Q_1(t)B_1^T(t) - P(t)C_1^T(t)R_1^{-1}(t)C_1(t)P(t), \quad t \in T_i, \\ P(t_{i-1}) = P_1(t_i) - P_1(t_i)C_2^T(t_i)[C_2(t_i)P_1(t_i)C_2^T(t_i) + R_2(t_i)]^{-1}C_2(t_i)P_1(t_i), \quad t_i \in \Theta, \\ P_1(t_i) = A_2(t_i)P(t_{i-1})A_2^T(t_i) + B_2(t_i)Q_2(t_i)B_2^T(t_i), \\ P(t_0) = P_0.\end{gathered} \tag{42}$$

**Алгоритм комплексирования.** В настоящей работе предлагается комплексирование дискретного и непрерывного каналов посредством решения непрерывной части уравнения для оценки (40) на каждом интервале дискретного времени $[t_i, t_{i+1}] \in \Theta$. При этом начальные условия решения $x(0) = x(t_i)$ равны оценке дискретного фильтра $\hat{x}(t_i)$, который предполагается более точным, но проигрывающим в частоте оценки.

**Расчеты.** Рассмотрим открытый контур системы управления погружением АНПА. Предполагается наличие дискретного и непрерывного измерителя глубины. Модель динамики и измерений непрерывной составляющей из уравнений (1) описывается матрицами

$$A_1 = \begin{bmatrix}0 & 1 \\ 0 & -0.5\end{bmatrix}, \qquad B_1 = \begin{bmatrix}0 \\ 0.01\end{bmatrix}, \qquad C_1 = [1 \quad 0].$$

Дискретная составляющая определяется иными матрицами состояния:

$$A_2 = \begin{bmatrix} 1 & 0.4424 \\ 0 & 0.7788 \end{bmatrix}, \qquad B_2 = \begin{bmatrix} 0.001152 \\ 0.004424 \end{bmatrix}, \qquad C_2 = [1 \quad 0].$$

В качестве эксперимента на входы $u_1$ и $u_2$ воздействует синусоидальный сигнал, интерпретируемый как сила воздействия на АНПА. Длительность процесса определяется множеством $T = [0{,}30]$, а моменты функционирования дискретной подсистемы $\Theta = \{0{,}0.5, \ldots, 30\}$. Вектор начального состояния при этом нулевой: $x(0) = [0 \quad 0]^T$, а матрица $P_0$ определяется из уравнения (15).

Шумовые воздействия $w_1$ и $v_1$ имеют идентичную ковариацию, а именно $Q_1 = R_1 = 0.01$. При этом, дискретный измеритель практически не подвержен влиянию шумов $w_2$ и $v_2$, значит $Q_2 = R_2 = 0$.

Результаты моделирования системы по уравнениям (40), (41) и (42) приведены на рис. 1, где построены ошибки оценки глубины АНПА.

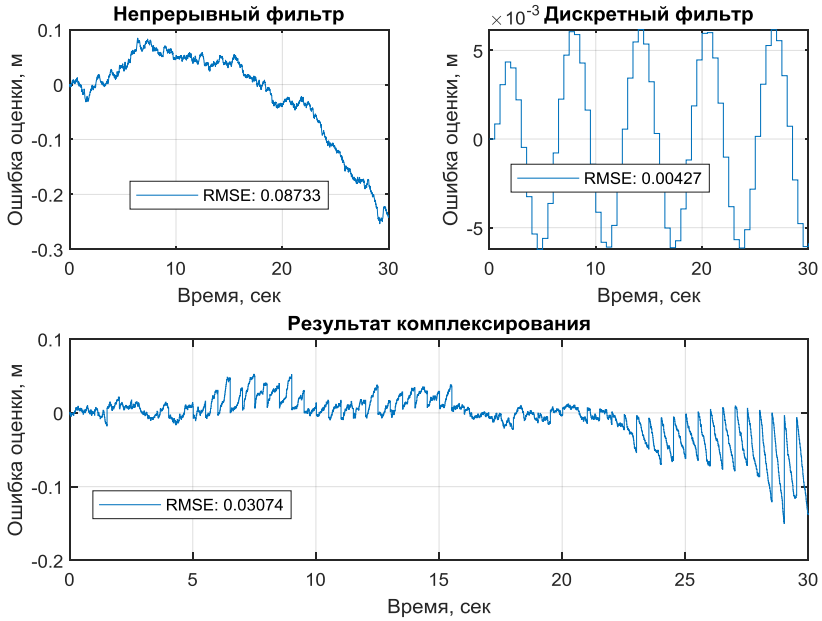



*Рис. 1. Результаты моделирования*

Графики ошибок относительно истинного состояния не зашумлённой системы для непрерывного и дискретного фильтров показывают, что последний значительно выигрывает в точности, однако не способен производить оценку состояния в реальном времени. Результат комплексирования показал, что среднеквадратическая ошибка (RMSE) в сравнении с непрерывным фильтром меньше на 35,2%, что подтверждает эффективность предложенного подхода.

**Заключение**. В результате проведенного исследования был разработан и обоснован алгоритм оптимального комплексирования оценки состояний в дискретно-непрерывных системах автономных необитаемых подводных аппаратов (АНПА). Предложенный подход демонстрирует высокую эффективность в повышении точности и надежности оценки параметров состояния АНПА путем оптимального объединения данных из различных источников информации.

Разработанный алгоритм обладает значительным потенциалом для применения в широком спектре областей, связанных с обработкой информации в сложных технических системах.

Важно отметить, что предложенный подход обладает гибкостью и масштабируемостью. Он может быть адаптирован для работы с большим количеством источников информации, что позволяет расширить его применимость в более сложных системах. Кроме того, алгоритм может быть модифицирован для использования в различных типах дискретно-непрерывных систем, не ограничиваясь только АНПА.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК


1. *Fossen T. I.* Handbook of marine craft hydrodynamics and motion control. – Hoboken, NJ Chichester, West Sussex: Wiley, 2021. – Second ed. – 710 c.
2. *Antonelli G.* Underwater Robots. – Cham: Springer International Publishing, 2018.
3. *Kramar V., Kabanov A., Dementiev K.* Autonomous Underwater Vehicle Navigation via Sensors Maximum-Ratio Combining in Absence of Bearing Angle Data // JMSE. – 2023. – Vol. 11, No. 10. – P. 1847.
4. *Щербатюк А.Ф.* О методе навигации группы АНПА без использования гидроакустических маяков // Гироскопия и навигация. – 2022. – Т. 30, № 4 (119). – С. 106-121.
5. *Jeong D. B., Ko N. Y.* Sensor Fusion for Underwater Vehicle Navigation Compensating Misalignment Using Lie Theory // Sensors. – 2024. – Vol. 24, No. 5. – P. 1653.
6. *Nicosevici T. et al.* A review of sensor fusion techniques for underwater vehicle navigation // Oceans ’04 MTS/IEEE Techno-Ocean ’04 (IEEE Cat. No.04CH37600). – Kobe, Japan: IEEE, 2004. – P. 1600-1605.
7. *Лазаренко А.Н.* Комплексирование ИНС и Ганс (GPS) фильтром Калмана // Вологдинские чтения. – 2008. – № 69.
8. *Вавилова Н.Б., Парусников Н.А., Субханкулова Г.А.* Навигация автономного подводного аппарата при помощи корректируемой бескарданной инерциальной навигационной системы // Тр. МАИ. – 2016. – № 89.
9. *Xincun Y. и др.* Kalman filter applied in underwater integrated navigation system // Geodesy and Geodynamics. – 2013. – Vol. 4, No. 1. – P. 46-50.
10. *Potokar E., Norman K., Mangelson J.* Invariant Extended Kalman Filtering for Underwater Navigation // IEEE Robot. Autom. Lett. – 2021. – Vol. 6, No. 3. – P. 5792-5799.
11. *Singh R.K., Saha J., Bhaumik S.* Maximum Correntropy Polynomial Chaos Kalman Filter for Underwater Navigation. – 2024.
12. *Sheng G. et al.* Cooperative Navigation Algorithm of Extended Kalman Filter Based on Combined Observation for AUVs // Remote Sensing. – 2023. – Vol. 15, No. 2. – P. 533.
13. *Allotta B. и др.* Sea currents estimation during AUV navigation using Unscented Kalman Filter // IFAC-PapersOnLine. – 2017. – Vol. 50, No. 1. – P. 13668-13673.
14. *Tijjani A.S. et al.* Continuous–Discrete Observation-Based Robust Tracking Control of Underwater Vehicles: Design, Stability Analysis, and Experiments // IEEE Trans. Contr. Syst. Technol. – 2023. – Vol. 31, No. 4. – P. 1477-1492.
15. *Andrienko A. Ya. et al.* On-Board Terminal Control Systems: Specifics and Design Concepts // IFAC Proceedings Volumes. – 1975. – Vol. 8, No. 1. – P. 464-471.
16. *Зорич В.А.* Математический анализ. Ч. II. – 6-е изд., дополн. – М.: МЦНМО, 2012. – 818 с.
17. *Барабанов А.Т.* Анализ и оптимизация функционально сложных систем гриновского типа на основе характеристического тождества // Адаптивные системы автоматического управления. – 1981. – № 9. – С. 3-12.
18. *Алексеев В.М., Тихомиров В.М., Фомин С.В.* Оптимальное управление. – М.: Наука, 1979. – 432 с.
19. *Острем К.Ю.* Введение в стохастическую теорию управления: пер. с англ. – М.: Мир, 1973. – 324 с.
20. *Casti J.* The linear-Quadratic Control Problem: Some Recent Results and Outstanding Problems // SIAM Rev. – 1980. – Vol. 22, No. 4. – P. 459-485.
21. *Зеликин М.И.* Оптимальное управление и вариационное исчисление. – 4-е изд., испр. – М.: ЛЕНАНД, 2017. – 160 с.


REFERENCES


1. *Fossen T.I.* Handbook of marine craft hydrodynamics and motion control. Hoboken, NJ Chichester, West Sussex: Wiley, 2021. Second ed., 710 p.
2. *Antonelli G.* Underwater Robots. Cham: Springer International Publishing, 2018.
3. *Kramar V., Kabanov A., Dementiev K.* Autonomous Underwater Vehicle Navigation via Sensors Maximum-Ratio Combining in Absence of Bearing Angle Data, *JMSE,* 2023, Vol. 11, No. 10, pp. 1847.

4. *Shcherbatyuk A.F.* O metode navigatsii gruppy ANPA bez ispol'zovaniya gidroakusticheskikh mayakov [On the method of navigation of a group of AUVs without the use of hydroacoustic beacons], *Giroskopiya i navigatsiya* [Gyroscopy and navigation], 2022, Vol. 30, No. 4 (119), pp. 106-121.
5. *Jeong D. B., Ko N. Y.* Sensor Fusion for Underwater Vehicle Navigation Compensating Misalignment Using Lie Theory, *Sensors,* 2024, Vol. 24, No. 5, pp. 1653.
6. *Nicosevici T. et al.* A review of sensor fusion techniques for underwater vehicle navigation, *Oceans '04 MTS/IEEE Techno-Ocean '04 (IEEE Cat. No.04CH37600)*. Kobe, Japan: IEEE, 2004, pp. 1600-1605.
7. *Lazarenko A.N.* Kompleksirovanie INS i Gans (GPS) fil'trom Kalmana [Integration of INS and HANS (GPS) with Kalman filter], *Vologdinskie chteniya* [Vologda readings], 2008, No. 69.
8. *Vavilova N.B., Parusnikov N.A., Subkhankulova G.A.* Navigatsiya avtonomnogo podvodnogo apparata pri pomoshchi korrektiruemoy beskardannoy inertsial'noy navigatsionnoy sistemy [Navigation of an autonomous underwater vehicle using a corrected strapdown inertial navigation system], *Tr. MAI* [Trudy MAI], 2016, No. 89.
9. *Xincun Y. u ∂p.* Kalman filter applied in underwater integrated navigation system, *Geodesy and Geodynamics,* 2013, Vol. 4, No. 1, pp. 46-50.
10. *Potokar E., Norman K., Mangelson J.* Invariant Extended Kalman Filtering for Underwater Navigation, *IEEE Robot. Autom. Lett*., 2021, Vol. 6, No. 3, pp. 5792-5799.
11. *Singh R.K., Saha J., Bhaumik S.* Maximum Correntropy Polynomial Chaos Kalman Filter for Underwater Navigation, 2024.
12. *Sheng G. et al.* Cooperative Navigation Algorithm of Extended Kalman Filter Based on Combined Observation for AUVs, *Remote Sensing,* 2023, Vol. 15, No. 2, pp. 533.
13. *Allotta B. u ∂p.* Sea currents estimation during AUV navigation using Unscented Kalman Filter, *IFAC-PapersOnLine,* 2017, Vol. 50, No. 1, pp. 13668-13673.
14. *Tijjani A.S. et al.* Continuous–Discrete Observation-Based Robust Tracking Control of Underwater Vehicles: Design, Stability Analysis, and Experiments, *IEEE Trans. Contr. Syst. Technol*., 2023, Vol. 31, No. 4, pp. 1477-1492.
15. *Andrienko A. Ya. et al.* On-Board Terminal Control Systems: Specifics and Design Concepts, *IFAC Proceedings Volumes,* 1975, Vol. 8, No. 1, pp. 464-471.
16. *Zorich V.A.* Matematicheskiy analiz [Mathematical analysis]. Part II. 6th ed. Moscow: MTSNMO, 2012, 818 p.
17. *Barabanov A.T.* Analiz i optimizatsiya funktsional'no slozhnykh sistem grinovskogo tipa na osnove kharakteristicheskogo tozhdestva [Analysis and optimization of functionally complex Green-type systems based on characteristic identity], *Adaptivnye sistemy avtomaticheskogo upravleniya* [Adaptive automatic control systems], 1981, No. 9, pp. 3-12.
18. *Alekseev V.M., Tikhomirov V.M., Fomin S.V.* Optimal'noe upravlenie [Optimal control]. Moscow: Nauka, 1979, 432 p.
19. *Ostrem K.Yu.* Vvedenie v stokhasticheskuyu teoriyu upravleniya [Introduction to stochastic control theory]: trans. from english. Moscow: Mir, 1973, 324 p.
20. *Casti J.* The linear-Quadratic Control Problem: Some Recent Results and Outstanding Problems, *SIAM Rev*., 1980, Vol. 22, No. 4, pp. 459-485.
21. *Zelikin M.I.* Optimal'noe upravlenie i variatsionnoe ischislenie [Optimal control and calculus of variations]. 4th ed. Moscow: LENAND, 2017, 160 p.

Статью рекомендовал к опубликованию д.т.н., профессор Ю.А. Кравченко.

**Кабанов Алексей Александрович** – Севастопольский государственный университет; e-mail: kabanovaleksey@gmail.com; г. Севастополь, Россия; тел.: +79787622582; кафедра «Информатика и управление в технических системах»; к.т.н.; доцент; зав. кафедрой.

**Крамарь Вадим Александрович** – e-mail: kramarv@mail.ru; тел.: +79787927340; кафедра «Информатика и управление в технических системах»; д.т.н.; профессор.

**Дементьев Кирилл Валерьевич** – e-mail: mash.saigon.89@gmail.com; тел.: +79788857479; кафедра «Информатика и управление в технических системах»; аспирант.

**Kabanov Aleksey Aleksandrovich** – Sevastopol State University; e-mail: kabanovaleksey@gmail.com; Sevastopol, Russia; phone: +79787622582; the Department of «Informatics and Control in Technical Systems»;cand. of eng. sc.; associate professor; department head.

**Kramar Vadim Aleksandrovich** – e-mail: kramarv@mail.ru; phone: +79787927340; the Department of «Informatics and Control in Technical Systems»; dr. of eng. sc.; professor.

**Dementiev Kirill Valerievich** – e-mail: mash.saigon.89@gmail.com; phone: +79788857479; the Department of "Informatics and Control in Technical Systems"; postgraduate student.

УДК 004.89 DOI 10.18522/2311-3103-2024-5-29-37

**В.В. Ковалев**

# АЛГОРИТМ ПРЕДВАРИТЕЛЬНОЙ ОБРАБОТКИ ИЗОБРАЖЕНИЙ ДЛЯ СНИЖЕНИЯ ВЕРОЯТНОСТИ ПЕРЕОБУЧЕНИЯ СВЁРТОЧНЫХ НЕЙРОННЫХ СЕТЕЙ НА НЕЙРОННОМ УСКОРИТЕЛЕ

*Основной объём требований в системах раннего обнаружения объектов предъявляется к производительности алгоритмов цифровой обработки изображений, которые реализуются на встраиваемых устройствах с ограниченным вычислительным ресурсом. В задаче раннего обнаружения объекты на изображениях представлены малым количеством пикселей. Поэтому чтобы обеспечить требуемые характеристики точности алгоритмов поиска и распознавания объектов на изображениях применяют алгоритмы предварительной обработки последовательности видеокадров для расширения исходного признакового пространства. Обработка изображений высокого разрешения алгоритмами предварительной обработки изображений приводит к неприемлемой временной задержке выполнения алгоритма и является «узким местом» всего алгоритма. В работе предложен алгоритм предварительной обработки последовательности видеокадров для нейронного ускорителя с целью расширения признакового пространства, который позволяет увеличить скорость обработки данных. Это достигается за счет слияний алгоритма предварительной обработки изображений с экстрактором признаков свёрточной нейронной сети и переносом выполнения нового экстрактора признаков на вычислительные мощности нейронного ускорителя. Произведена апробация разработанного алгоритма путём проведения вычислительного эксперимента. На вычислительных устройствах NVIDIA Jetson и Rockchip реализован алгоритм предварительной обработки дважды на центральном процессоре и нейронном ускорителе, согласно разработанному алгоритму. Получены оценки времени выполнения алгоритмов, которые показывают, что предложенный алгоритм предварительной обработки изображений для нейронного ускорителя позволяет увеличить скорость обработки данных в 1.4–4 раз в зависимости от типа разрядности вычислений. Однако, переход к целочисленному типу вычислений модели СНС с модифицированным экстрактором признаков приводит к снижению метрики Mean Average Precision на 5–19.4%, характеризующей интегральную среднюю точность поиска и распознавания объектов на изображениях.*

*Свёрточные нейронные сети; вычислительные устройства; распознавание образов; вычислительные устройства.*

**V.V. Kovalev**

# IMAGE PREPROCESSING ALGORITHM TO REDUCE THE PROBABILITY OF OVERFITTING OF CONVOLUTIONAL NEURAL NETWORKS ON A NEURAL ACCELERATOR

*The main volume of requirements in early detection systems are imposed on the performance of digital image processing algorithms that are implemented on embedded devices with limited computing resources. In the early detection problem, objects in images are represented by a small number of pixels. Therefore, in order to ensure the required accuracy characteristics of the algorithms for searching and recognizing objects in images, algorithms for preliminary processing of a sequence of video frames are used to expand the original feature space. Processing high-resolution images by image preliminary processing algorithms leads to an unacceptable time delay in the execution of the algorithm and is a "bottleneck" of the entire algorithm. In this paper, an algorithm for preliminary processing of a sequence of video frames for a neural accelerator is proposed in order to expand the feature space, which allows increasing the speed of data processing. This is achieved by merging the image preliminary processing algorithm with the feature extractor of a convolutional neural network and transferring the execution of a new feature extractor to the computing power of the neural accelerator. The developed algorithm was tested by conducting a computational experiment. On NVIDIA Jetson and Rockchip computing devices, the algorithm of preliminary processing is implemented twice on the central processor and neural accelerator, according to the developed algorithm. Estimates of the execution time of the algorithms are obtained, which show that the proposed algorithm of preliminary image processing for the neural accelerator al-*

*lows increasing the data processing speed by 1.4 - 4 times depending on the type of bit depth of calculations. However, the transition to the integer type of calculations of the CNN model with a modified feature extractor leads to a decrease in the Mean Average Precision metric by 5–19.4%, characterizing the integral average accuracy of searching and recognizing objects in images.*

*Convolutional neural networks; computing devices; pattern recognition; computing devices.*

**Введение.** Поиск и распознавание объектов на изображениях является неотъемлемой задачей современных систем анализа и обработки видеопоследовательности. Такие системы задают высокие требования к критерию точность/быстродействие, чтобы своевременно и точно принимать решения во избежание аварийных ситуаций. Для обеспечения высокой точности поиска и распознавания объектов на изображениях в масштабе реального времени используют свёрточные нейронные сети (СНС). Чтобы обеспечить своевременное принятие решения, бортовые системы должны обнаруживать и распознавать объекты на дальних расстояниях. Чтобы изображение объекта на дальнем расстоянии от места фиксации имело не нулевое признаковое описание, физический размер объектов должен быть большим. Но всё же такие объекты представлены малой группой пикселей на изображении относительно всей площади изображения. Поиск и распознавание малоразмерных объектов на изображениях вызывает трудности даже у передовых архитектур СНС. Неудовлетворительные показатели точности поиска и распознавания малоразмерных объектов связаны с проблемой переобучения СНС. Переобучение – это негативный процесс, который возникает когда объем признакового пространства мал относительно количества степеней свободы (обучаемых параметров) СНС. Характерным проявлением переобучения модели СНС является достижение высоких показателей точности поиска и распознавания объектов на данных, которые были в обучающей выборке, а на тестовых данных, которые сеть «не видела» алгоритм показывает низкую точность поиска и распознавания. Для борьбы с переобучение применяют специальные методы, которые объединил термин регуляризация. Методы регуляризации условно можно разделить на три группы: изменение функции потерь, изменение процесса обучения и архитектуры СНС, преобразование данных. В последнее время большую популярность набирают методы преобразования данных, которые основаны на расширении признакового пространства, которые основываются на алгоритмах предварительной обработки. Однако, алгоритмы предварительной обработки в поиске и распознавание малоразмерных объектов СНС могут быть вычислительно затратными и является «узким местом» выполнения алгоритма в целом, потому что построены не центральном процессоре. Поэтому, проблемы неудовлетворительной временной задержке алгоритмов предварительной обработки изображений является актуальной задачей, которая требует решения.

**Поиск и распознавание объектов на изображениях свёрточными нейронными сетями.** СНС поиска и распознавания объектов на изображениях условно можно разделить на двухэтапные и одноэтапные подходы. Двухэтапные походы осуществляют поиск и распознавание объектов в два этапа: 1) формируют множество потенциальных регионов; 2) классификация каждого региона. Одноэтажные подходы объединяют эти два этапа в один и напрямую обнаруживают и распознают объекты. В свою очередь двухэтапные подходы обеспечивают большую точность поиска и распознавания по сравнению с одноэтапными подходами. Однако, «платой» за точность поиска и распознавания является большая вычислительная сложность алгоритма.

Поэтому в распознающих системах для автономных носителей [1] используются одноэтапные СНС, которые могут обрабатывать данные в масштабе реального времени с высокой точностью распознавания на устройствах с ограниченным вычислительным ресурсом.

Входными данными для СНС могут быть изображения, последовательность изображений (видеокадров), n-мерные тензоры (батчи), пирамиды изображений. Современные одноэтапные СНС состоят из экстрактора признаков (Backbone) и блока декодирования (Head) (рис. 1). Самой вычислительно затратной составной частью алгоритма является экстрактор признаков, поэтому прямое распространение данных через экстрактор при-

знаков осуществляется на нейронном ускорителе. К числу нейронных ускорителей относятся такие устройства как: графический процессор или Graphics Processing Unit (GPU); нейронный процессор или Neural Processing Unit (NPU); тензорный процессор или Tensor Processing Unit (TPU); программируемая логическая интегральная схема (ПЛИС) или Field Programmable Gate Array (FPGA) и др.

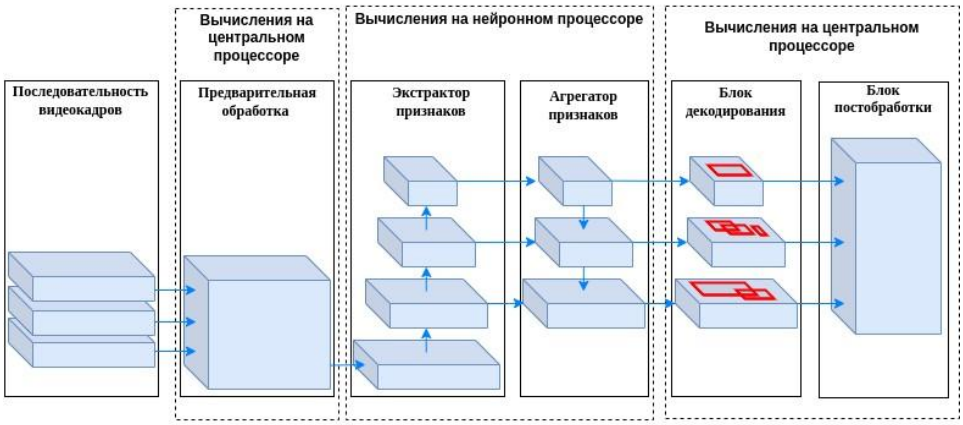



*Рис. 1. Структурная схема алгоритма поиска и распознавания объектов на изображениях для устройств с ограниченным вычислительным ресурсом*

Современные архитектуры СНС используют промежуточные слои между экстрактором признаков и блоком декодирования для того, чтобы собрать семантические информацию с различных масштабов пирамиды признаков. Такое слияние признаков называют агрегированием или Neck. Агрегирование признаков может быть вычислительно затратным, поэтому этот этап алгоритма тоже выполняется на нейронном ускорителе. Декодирование агрегированных признаков осуществляется одноэтапными алгоритмами таким как: YOLO [2–5], SSD [6], CenterNet [7], CornerNet [8], FCOS [9] и др. Одной из проблем алгоритмов поиска и распознавания объектов является предсказание двойных обрамляющих прямоугольников для одного объекта на изображении. Для устранения данной проблем применяют алгоритм постобработки Non Maximum Suppression (NMS) [10]. Декодирование и постобработка не являются вычислительно затратными этапами алгоритма, поэтому выполняется на центральном процессоре.

Поиск и распознавания малоразмерных объектов на изображениях вызывает трудности даже у передовых архитектур СНС из-за проблемы переобучения. Поэтому применяют методы регуляризации, которые направлены на снижение вероятности переобучения СНС за счёт расширения исходного признакового пространства. Обычно, изображения видимого оптического диапазона расширяют данными различных типов (мультимодальные изображения): инфракрасные (ИК) изображения, данные с лазерных дальномеров, радиолокационные изображения, изображения с признаком движения и др. Расширение признакового пространства радиолокационными изображениями, изображениями с лазерных дальномеров и ИК изображениями имеет существенные недостатки, заключающиеся в применении дорогостоящей сенсорной аппаратуры и усложнении конструкции распознающей системы. В свою очередь, изображения с признаками движения можно получить алгоритмами предварительной обработки последовательности изображений [11–14]. Однако, обработка изображений высокого разрешения на центральном процессоре может приводить к неприемлемой временной задержке и являться «узким местом» всего алгоритма. Поэтому предложен алгоритм предварительной обработки последовательности изображений на нейронном ускорителе для снижения временной задержки.

**Алгоритм предварительной обработки последовательности изображений для расширения признакового пространства.** Алгоритм предварительной обработки последовательности видеокадров направлен на расширения исходного признакового про-

странства признаками движения, полученными в процессе длительного наблюдения за объектом. На вход алгоритма поступает последовательность изображений, которая накапливается в линии задержке с отводами. Период выдачи кадров камерой определяется временной задержкой $T$. Накопленная последовательность видеокадров поступает в блок идентификации движения. С учётом компромисса между точностью и вычислительной сложностью, алгоритм идентификации движения основан на построении разностных изображений между опорным изображением и задержанными изображениями [15]:

$$d_i(x,y) = |f(x,y,t_0) - f(x,y,t_{i+1})|, \quad (1)$$

где $i$ есть индекс разностного изображения, который находится в диапазон от $0$ до $k-1$, где $k$ длина очереди видеокадров; $f(x,y,t_0)$ и $f(x,y,t_{i+1})$ – дискретные изображения, полученные в моменты времени $t_0$, $t_{i+1}$; $x$, $y$ – пространственные координаты. Изображение, полученное в момент времени $t_0$, является опорным изображением, а в момент времени $t_{i+1}$ является задержанным. Разностные изображения $d_i$ формируются путем сравнения опорного изображения с задержанным изображением. Таким образом, разностное изображение, полученное в начальные моменты времени, отображает наличие медленного движения в видеокадре, а в позднее моменты времени отображает быстрое наличие движение. Взятие абсолютных значений разностных изображений позволяет привести значения яркостей изображения к одному динамическому диапазону от нуля до единицы во всех каналах комплексированного изображения.

Результирующее псевдоцветное изображение формируется на основе поканального объединения на уровне пикселей опорного изображения и $k$ разностных изображений для заданных моментов времени. Таким образом, количество каналов разностных изображений и время задержки кадров не фиксированы и задаются разработчиком. Конечным результатом разработанного алгоритма предварительно обработки изображений является комплексированное изображение характерное «цветными» локальными областями, означающими признак движения.

Алгоритм предварительной обработки последовательности изображений для расширения признакового пространства на нейронном ускорителе. Концепция построения алгоритмов СНС на устройствах с ограниченным вычислительным ресурсом заключается в переносе вычислительно сложных этапов алгоритма с центрального процессора на нейронный ускоритель. Это достигается за счет объединения вычислений предварительной обработки изображений с экстрактором признаков СНС и выполнения на нейронном ускорителе. Структурная схема алгоритма поиска и распознавания объектов на изображениях с учётом применения алгоритма предварительной обработки последовательности изображений на нейронном ускорителе измениться согласно структурной схеме, представленной на рис. 2.

Согласно структурной схеме представленной на рис. 2, необходимо объединить последовательность операций, которые включает алгоритм предварительной обработки с последовательностью операций экстрактора признаков. Для реализации данного похода необходимо представить вычисления алгоритма предварительной обработки в рамках операций поддерживаемых нейронными ускорителями. К числу операций, которые включает алгоритм предварительной обработки изображений для расширения признакового пространства, относятся: поэлементная (попиксельная) матричная разность и вычисление абсолютных значений каждого элемента матрицы.

Алгоритм использует опорное изображение и два изображения, задержанных на $T$ и $4T$. С учётом выбранных задержек, для получения комплексированного изображения необходимо накопить последовательность из пяти изображений. На основе накопленной последовательности формируется тензор из изображений, полученных в настоящий момент времени, $T$ и $4T$. Формирование опорного и двух разносных изображений реализовано с помощью операции двумерная свёртка (2D Convolution) размерностью фильтров $3\ x\ 3\ x\ 1\ x\ 1$, со значениями $0, 1, -1$. Единичные значения фильтров формируют опорное изображение, отрицательные значения формируют изображение с отрицательными значениями, нулевые значения удаляют все признаки изображения. Фильтр с очередностью

значений $1, 0, 0$ формирует опорное изображение. Очередность значений ядер фильтра $1, -1, 0$ формирует разностное изображение, характеризующее медленное наличие движения, а фильтр с значениями $1, 0, -1$ формирует разностное изображение, характеризующее быстрое наличие движения. Таким образом, на выходе свёртки формируется тензор равный по размерности входному тензору.

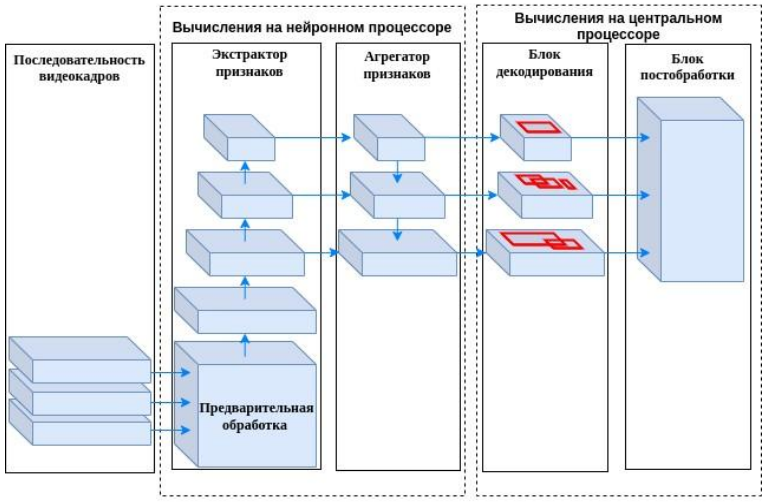



*Рис. 2. Структурная схема алгоритма поиска и распознавания объектов с учётом применения алгоритма предварительной обработки последовательности изображений на нейронном ускорителе*

Опорное изображение имеет только положительные значения, а значения в разностных каналах могут иметь отрицательные значения. Чтобы привести значения выходного тензора к одному динамическому диапазону от 0 до 1 во всех каналах определяются абсолютные значения всех элементов тензора. Результатом всех операций является тензор, состоящего из опорного и двух разностных изображений. Если представить тензор в цветовом пространстве Red Green Blue (RGB), то опорное изображение находится в цветовом канале Red, а разностные изображения в цветовых каналах Green и Blue.

Экспериментальное исследование алгоритма предварительной обработки последовательности изображений для расширения признакового пространства на нейронном ускорителе. В экспериментальной части данной главы произведено сравнение эффективности предложенного алгоритма путем сравнения времени выполнения алгоритма на нейронном ускорители и центральном процессоре. Для проведения эксперимента сформировано множество вычислительных устройств с аппаратной поддержкой нейронного ускорителя: NVIDIA Jetson Nano, NVIDIA Jetson TX2, NVIDIA Jetson AGX Xavier и Rockchip RK3588.

*Построение алгоритма на вычислительных устройствах NVIDIA Jetson.*

Построение алгоритма на вычислительных устройствах NVIDIA Jetson. Для построения алгоритма комплексирования информации о движении на устройствах линейки Jetson используется среда разработки JetPack. Для выполнения алгоритма расширения признакового пространства на центральном процессоре используется фреймворк Pytorch, который реализует CPU-модель алгоритма. GPU-модели алгоритма построены с помощью фреймворка TensorRT.

*Построение алгоритма на вычислительном устройстве Rickchip.*

В необходимый набор программных средств для построения NPU-модели алгоритма предварительной обработки изображений для расширения признакового пространства на вычислительном устройстве Rockchip RK3588 входят ОС Linux Ubuntu SDK (Software Development Kit) Rockchip Neural Network-Toolkit2 (RKNN-Toolkit2) и RKNN-Toolkit-

Lite2. RKNN-Toolkit2 — это комплект средств разработки программного обеспечения, позволяющий выполнять преобразование моделей, прямой вывод и оценку производительности на персональных компьютерах и вычислительных устройствах Rockchip. RKNN-Toolkit-Lite2 предоставляет интерфейсы программирования Python для платформы Rockchip, помогающие строить NPU-модели и ускорять выполнение.

Оценка времени выполнения алгоритма предварительной обработки изображений на устройствах производилась в режиме максимально допустимой вычислительной мощности для всех устройств: устройства линейки Jetson в режиме *jetson_clocks*, устройства Rockchip RK3588 режим загрузки всех вычислительных ядер нейронного процессора.

Оценка времени выполнения Pytorch CPU-моделей алгоритма предварительной обработки изображений для всех вычислительных устройств оценивается с помощью профилировщика profile входящего в фреймворк Pytorch.

На каждом вычислительном устройстве построены модели алгоритма предварительной обработки изображений для расширения признакового пространства для всех поддерживаемых типов разрядности вычислений. После чего произведена оценка времени выполнения построенных моделей алгоритма для размера входных данных $1\ x\ 3\ x\ 960\ x\ 960$ элементов, которые отображены в табл. 1. Переход к вычислениям с пониженной разрядностью приводит к снижению точности поиска и распознавания объектов на изображениях [16, 17]. Поэтому произведём вычислительный эксперимент влияния перехода к вычислениям с пониженной разрядностью.

Для оценки точности поиска и распознавания малоразмерных объектов сформирован датасет синтетических изображений с аннотациями с помощью средств виртуального моделирования 3D графика Unreal Engine 5. Датасет состоит из аннотированных условно-реальных изображений и включает 5 классов техники. Датасет сформирован из 80 видео сюжетов видимого оптического диапазона различной местности. Из первых 50 сюжетов получена тренировочная выборка, а из остальных 30 тестовая выборка. В тренировочную выборку входят 50000, а в тестовую 30000 аннотированных изображений. На изображении присутствуют не менее одного малоразмерного объекта.

Таблица 1

**Оценка времени выполнения моделей алгоритма предварительной обработки изображений для расширения признакового пространства на вычислительных устройствах**

| Вычислительное устройство | Нейронный ускоритель | Точность вычислений | Время выполнения, мс |
|---|---|---|---|
| NVIDIA Jetson Nano B01 | CPU | FP32 | 26.1 |
| | GPU | FP32 | 18.2 |
| | | FP16 | 12.4 |
| NVIDIA Jetson TX2 8GB | CPU | FP32 | 23.5 |
| | GPU | FP32 | 10.9 |
| | | FP16 | 7.4 |
| NVIDIA Jetson AGX Xavier | CPU | FP32 | 8.4 |
| | GPU | FP32 | 3.1 |
| | | FP16 | 2.5 |
| | | INT8 | 2.1 |
| Rockckip RK3588 | CPU | FP32 | 16.4 |
| | NPU | FP32 | 15.9 |
| | | FP16 | 11.2 |

Анализ научных работ в задаче обнаружения малоразмерных объектов показал, что нет общепринятого критерия малоразмерности объектов. В исследовании [18] авторы выбрали критерий малоразмерности для объектов, физический размер которых не превышал 0.3 м, а относительное перекрытие площади обрамляющего прямоугольника объекта с площадью изображения находилось в пределах от 0.08 до 0.58. В вычислительном эксперименте используется это условие только для объектов с большим физическим размером (до 6 м).

Произведена оценка точности поиска и распознавания малоразмерных объектов моделей СНС на тестовом датасете на основе метрики Mean Average Precision (mAP) [19, 20]. Метрики mAP получены для входных данных изображений размерностью 3 x 960 x 960 и представлены в табл. 2. Исходя из полученных оценок времени выполнения, можно сделать, что предложенный алгоритм позволяет увеличить скорость обработки данных предварительной обработки в 1.4–4 раз в зависимости типа нейронного ускоритель и разрядности вычислений (битности). Однако, переход к вычислениям с целочисленным типом разрядности (integer 8) приводит к снижению интегрального показателя точности поиска и распознавания Mean Average Precision на 5-19.4%.

Таблица 2

**Оценка точности поиска и распознавания малоразмерных объектов СНС для вычислений с различной точностью**

| Архитектура СНС | Нейронный ускоритель | Точность вычислений | mAP, % |
|---|---|---|---|
| YOLOv3-Tiny-M | GPU | FP32 | 66.67 |
| | | FP16 | 66.79 |
| | | INT8 | 49.732 |
| | NPU | FP32 | 66.67 |
| | | INT8 | 52.8 |
| YOLOv5s-M | GPU | FP32 | 73.29 |
| | | FP16 | 73.29 |
| | | INT8 | 54.15 |
| | NPU | FP32 | 73.29 |
| | | INT8 | 63.4 |
| YOLOv5l-M | GPU | FP32 | 86.77 |
| | | FP16 | 86.77 |
| | | INT8 | 80.3 |
| | NPU | FP32 | 86.77 |
| | | INT8 | 81.7 |
| YOLOv8s-M | GPU | FP32 | 73.27 |
| | | FP16 | 73.3 |
| | | INT8 | 65.1 |
| | NPU | FP32 | 73.27 |
| | | INT8 | 64.3 |

**Заключение.** Произведено экспериментальное исследование времени выполнения алгоритма предварительной обработки последовательности видеокадров для расширения признакового пространства на нейронном процессоре. На основе полученных оценок

времени выполнения алгоритма можно сделать выводы, что применение предложенного подхода позволило сократить время выполнения алгоритма в 1.4-4 раз в зависимости от типа нейронного ускорителя и разрядности (битности) вычислений по сравнению с центральным процессором. Переход к вычислениям с половинной разрядностью (float pointing 16) не изменяет точность поиска и распознавания объектов, а переход к целочисленной разрядности вычислений (integer 8) приводит к снижению критерия Mean Average Precision на 5-19.4 %, характеризующего интегральную среднюю точность.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК


1. *Алпатов Б. и др.* Методы автоматического обнаружения и сопровождения объектов. Обработка изображений и управление. – М.: Радиотехника, 2008. – 176 с.
2. *Redmon J., Farhadi A.* YOLO9000: better, faster, stronger // Proceedings of the IEEE conference on computer vision and pattern recognition. – 2017. – P. 7263-7271.
3. *Redmon J., Farhadi A.* Yolov3: An incremental improvement // arXiv preprint arXiv:1804.02767. – 2018.
4. *Bochkovskiy A., Wang C.-Y., Liao H.-Y.M.* Yolov4: Optimal speed and accuracy of object detection // arXiv preprint arXiv:2004.10934. – 2020.
5. *Wang C.-Y., Bochkovskiy A., Liao H.-Y.M.* Scaled-yolov4: Scaling cross stage partial network // Proceedings of the IEEE/cvf conference on computer vision and pattern recognition. – 2021. – P. 13029-13038.
6. *Liu W., et al.* SSD: Single shot multibox detector // Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14. – Springer, 2016. – P. 21-37.
7. *Duan K., et al.* Centernet: Keypoint triplets for object detection // Proceedings of the IEEE/CVF international conference on computer vision. – 2019. – P. 6569-6578.
8. *Law H., Deng J.* Cornernet: Detecting objects as paired keypoints // Proceedings of the European conference on computer vision (ECCV). – 2018. – P. 734-750.
9. *Tian Z., et al.* Fcos: Fully convolutional one-stage object detection // Proceedings of the IEEE/CVF international conference on computer vision. – 2019. – P. 9627-9636.
10. *Girshick R., et al.* Rich feature hierarchies for accurate object detection and semantic segmentation // Proceedings of the IEEE conference on computer vision and pattern recognition. – 2014. – P. 580-587.
11. *Niu Y., et al.* Airborne infrared and visible image fusion for target perception based on target region segmentation and discrete wavelet transform // Mathematical Problems in Engineering. – 2012. – T. 2012.
12. *Liu S., Liu Z.* Multi-channel CNN-based object detection for enhanced situation awareness // arXiv preprint arXiv:1712.00075. – 2017.
13. *Smeelen M.A., et al.* Semi-hidden target recognition in gated viewer images fused with thermal IR images // Information Fusion. – 2014. – Vol. 18. – P. 131-147.
14. *Li Y., et al.* Multimodal medical supervised image fusion method by CNN // Frontiers in Neuroscience. – 2021. – Vol. 15. – P. 638976.
15. *Ковалев В.В.* Методы решения проблемы переобучения нейронных сетей в задаче обнаружения малоразмерных объектов на изображениях // Тр. международного научно-технического конгресса «Интеллектуальные системы и информационные технологии-2023». – 2023. – P. 38-45.
16. *Ковалев В.В., Сергеев Н.Е.* Реализация сверточных нейронных сетей на встраиваемых устройствах с ограниченным вычислительным ресурсом // Известия ЮФУ. Технические науки. – 2021. – № 6 (223). – С. 64-72.
17. *Щелкунов А.Е. и др.* Ускорение прямого прохода при реализации СНС на ограниченном вычислительном ресурсе // Известия ЮФУ. Технические науки. – 2022. – № 1. – С. 289-297.
18. *Chen C., et al.* R-CNN for small object detection // Computer Vision–ACCV 2016: 13th Asian Conference on Computer Vision, Taipei, Taiwan, November 20-24, 2016, Revised Selected Papers, Part V 13. – Springer, 2017. – P. 214-230.
19. *Everingham M., et al.* The pascal visual object classes (voc) challenge // International journal of computer vision. – 2010. – Vol. 88. – P. 303-338.
20. *Padilla R., Netto S.L., Da Silva E.A.* A survey on performance metrics for object-detection algorithms // 2020 international conference on systems, signals and image processing (IWSSIP). – IEEE, 2020. – P. 237-242.

REFERENCES


1. *Alpatov B. i dr.* Metody avtomaticheskogo obnaruzheniya i soprovozhdeniya ob"ektov. Obrabotka izobrazheniy i upravlenie [Methods of automatic detection and tracking of objects. Image processing and control]. Moscow: Radiotekhnika, 2008, 176 p.
2. *Redmon J., Farhadi A.* YOLO9000: better, faster, stronger, *Proceedings of the IEEE conference on computer vision and pattern recognition,* 2017, pp. 7263-7271.
3. *Redmon J., Farhadi A.* Yolov3: An incremental improvement, *arXiv preprint arXiv:1804.02767,* 2018.
4. *Bochkovskiy A., Wang C.-Y., Liao H.-Y.M.* Yolov4: Optimal speed and accuracy of object detection, *arXiv preprint arXiv:2004.10934,* 2020.
5. *Wang C.-Y., Bochkovskiy A., Liao H.-Y.M.* Scaled-yolov4: Scaling cross stage partial network, *Proceedings of the IEEE/cvf conference on computer vision and pattern recognition,* 2021, pp. 13029-13038.
6. *Liu W., et al.* SSD: Single shot multibox detector, *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14*. Springer, 2016, pp. 21-37.
7. *Duan K., et al.* Centernet: Keypoint triplets for object detection, *Proceedings of the IEEE/CVF international conference on computer vision,* 2019, pp. 6569-6578.
8. *Law H., Deng J.* Cornernet: Detecting objects as paired keypoints, *Proceedings of the European conference on computer vision (ECCV),* 2018, pp. 734-750.
9. *Tian Z., et al.* Fcos: Fully convolutional one-stage object detection, *Proceedings of the IEEE/CVF international conference on computer vision,* 2019, pp. 9627-9636.
10. *Girshick R., et al.* Rich feature hierarchies for accurate object detection and semantic segmentation, *Proceedings of the IEEE conference on computer vision and pattern recognition,* 2014, pp. 580-587.
11. *Niu Y., et al.* Airborne infrared and visible image fusion for target perception based on target region segmentation and discrete wavelet transform, *Mathematical Problems in Engineering,* 2012, T. 2012.
12. *Liu S., Liu Z.* Multi-channel CNN-based object detection for enhanced situation awareness, *arXiv preprint arXiv:1712.00075,* 2017.
13. *Smeelen M.A., et al.* Semi-hidden target recognition in gated viewer images fused with thermal IR images, *Information Fusion,* 2014, Vol. 18, pp. 131-147.
14. *Li Y., et al.* Multimodal medical supervised image fusion method by CNN, *Frontiers in Neuroscience,* 2021, Vol. 15, pp. 638976.
15. *Kovalev V.V.* Metody resheniya problemy pereobucheniya neyronnykh setey v zadache obnaruzheniya malorazmernykh ob"ektov na izobrazheniyakh [Methods for solving the problem of overtraining neural networks in the task of detecting small-sized objects in images], *Tr. mezhdunarodnogo nauchno-tekhnicheskogo kongressa «Intellektual'nye sistemy i informatsionnye tekhnologii-2023»* [Proceedings of the international scientific and technical congress "Intelligent systems and information technologies-2023"], 2023, pp. 38-45.
16. *Kovalev V.V., Sergeev N.E.* Realizatsiya svertochnykh neyronnykh setey na vstraivaemykh ustroystvakh s ogranichennym vychislitel'nym resursom [Implementation of convolutional neural networks on embedded devices with limited computing resources], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2021, No. 6 (223), pp. 64-72.
17. *Shchelkunov A.E. i dr.* Uskorenie pryamogo prokhoda pri realizatsii SNS na ogranichennom vychislitel'nom resurse [Acceleration of the forward pass when implementing a CNN on a limited computing resource] *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2022, No. 1, pp. 289-297.
18. *Chen C., et al.* R-CNN for small object detection, *Computer Vision–ACCV 2016: 13th Asian Conference on Computer Vision, Taipei, Taiwan, November 20-24, 2016, Revised Selected Papers, Part V 13*. Springer, 2017, pp. 214-230.
19. *Everingham M., et al.* The pascal visual object classes (voc) challenge, *International journal of computer vision,* 2010, Vol. 88, pp. 303-338.
20. *Padilla R., Netto S.L., Da Silva E.A.* A survey on performance metrics for object-detection algorithms, *2020 international conference on systems, signals and image processing (IWSSIP)*. IEEE, 2020, pp. 237-242.


Статью рекомендовал к опубликованию д.т.н., профессор Ю.А. Кравченко.

**Ковалев Владислав Владимирович** – Южный федеральный университет; e-mail: vlad.kovalev94@mail.ru; г. Таганрог, Россия; тел.: +79525864492; кафедра вычислительной техники; соискатель.

**Kovalev Vladislav Vladimirovich** – Southern Federal University**;** e-mail: vlad.kovalev94@mail.ru; Taganrog, Russia; phone: +79525864492; the Department of Computer Science; post-graduate student.

УДК 004.056 DOI 10.18522/2311-3103-2024-5-38-47

**И.В. Котенко, М.В. Мельник**

# ПРИМЕНЕНИЕ ГИБРИДНОЙ НЕЙРОННОЙ СЕТИ AE-LSTM ДЛЯ ОБНАРУЖЕНИЯ АНОМАЛИЙ В КОНТЕЙНЕРНЫХ СИСТЕМАХ

*Популярность контейнерных систем привлекает внимание многих исследователей в области информационных технологий. Технология контейнеризации позволяет сократить расходы вычислительных ресурсов при разворачивании и поддержке сложных инфраструктурных решений. Обеспечение безопасности контейнерных систем и контейнеризации в целом, а также применение злоумышленниками методов реализации "умных" атак на основе искусственного интеллекта, является серьезной проблемой на пути безопасного и устойчивого функционирования контейнерных систем. В статье предлагается подход к обнаружению не только ранее неизвестных отдельных аномальных процессов, но и аномальных последовательностей процессов в контейнерных системах. Предлагаемый подход и его реализация на основе платформы Docker основываются на трассировке системных вызовов, построении гистограмм выполняемых процессов и использовании нейронной сети AE-LSTM. Процесс построения гистограмм базируется на учете количества выполненных системных вызовов для каждого отдельного процесса. Это решение предоставляет возможность не только идентифицировать любой процесс в системе, но и эффективно обнаруживать аномальные последовательности процессов с высокой степенью точности. Созданные последовательности используются в качестве входных данных для нейронной сети. После завершения процесса обучения, нейронная сеть приобретает способность обнаруживать аномальные последовательности, сравнивая заданный порог ошибки реконструкции с фактическим уровнем ошибки входного вектора данных. Когда нейронная сеть сталкивается с новым входным вектором данных, она вычисляет уровень ошибки реконструкции — разницу между ожидаемым и фактическим значением. Если эта ошибка превышает заранее установленный порог, система сигнализирует о наличии аномалии в последовательности. Эксперименты показывают, что предложенный подход демонстрирует достаточно высокую точность обнаружения аномальных процессов при низком уровне ложноположительных результатов обнаружения. Такие результаты подтверждают эффективность предложенного подхода. Затраты вычислительных ресурсов на обучение модели нейронной сети находятся на достаточно низком уровне. Это позволяет использовать менее мощные аппаратные средства без значительных потерь в производительности. Разработанный прототип может быть обучен и внедрен в новую инфраструктуру в достаточно сжатые сроки.*

*Обнаружение аномалий; системные вызовы; контейнерные системы; нейронные сети; автоэнкодер; долговременная кратковременная память.*

**I.V. Kotenko, M.V. Melnik**

# APPLICATION OF A HYBRID NEURAL NETWORK AE-LSTM FOR ANOMALIES DETECTION IN CONTAINER SYSTEMS

*The popularity of container systems attracts the attention of many researchers in the field of information technology. Containerization technology allows to reduce the cost of computing resources when deploying and supporting complex infrastructure solutions. Ensuring the security of container systems and containerization in general, as well as the use of smart attacks based on artificial intelligence by malefactors, is a serious problem on the way to the safe and stable operation of container systems. This article proposes an approach for detecting not only previously unknown individual anomalous processes, but also anomalous process sequences in container systems. The proposed approach and its implementation based on the Docker platform are based on tracing system calls, constructing histograms of running processes, and using the AE-LSTM neural network. The process of constructing histograms is based on accounting of the number of executed system calls for each individual process. This solution provides the ability not only to accurately identify any process in the system, but also to effectively detect anomalous process sequences with a high degree of accuracy. The generated sequences are used as input data for the neural network. After completing the training process, the neural network acquires the ability to detect anomalous sequences by comparing a given threshold of reconstruction error with the actual error level of the input data vector. When the neural network encounters a new input data vector, it calculates the reconstruction error level - the difference between the expected and actual value. If this error exceeds a predetermined*

*threshold, the system signals the presence of an anomaly in the sequence. Experiments show that the proposed approach demonstrates high accuracy in detecting anomalous processes with a low level of false positive detection results. Such results confirm the effectiveness of the proposed approach. Also, the computational costs of training the neural network model are quite low. This allows using less powerful hardware without significant performance losses. Such a solution can be trained and implemented in a new infrastructure in a fairly short time.*

*Anomaly detection; system calls; container systems; neural networks; autoencoder; long short-term memory.*

**Введение.** В последние годы, использование контейнерных систем позволило повысить удобство в эксплуатации и масштабируемость, а также снизить затраты вычислительных ресурсов, и в настоящее время широко используются при внедрении сложных и высоконагруженных вычислительных систем. Такие преимущества привлекают разработчиков программного обеспечения. Но проблемы безопасности и активное применение методов искусственного интеллекта в проведении атак [1], является серьезным препятствием на пути внедрения технологии контейнеризации.

Только в 2024 году в системах виртуализации было обнаружено несколько серьезных уязвимостей, например, CVE-2024-21626, CVE-2024-23651, CVE-2024-23652 и CVE-2024-23653 [2]. Данные уязвимости позволяют злоумышленнику покинуть пределы изолированных контейнерных систем и получить несанкционированный доступ к данным хостовой операционной системы. Также дополнительные проблем несут возможные нарушения конфигурации, уязвимые образы, пренебрежение механизмами изоляции контейнерных систем и запуск контейнеров с правами администратора (root) [3].

Киберпреступники часто усовершенствуют свои инструменты, тактики и техники, используемые при атаках на информационные системы, что ставит под сомнение использование средств защиты на основе сигнатурного подхода [4]. Решения на основе анализа аномалий [5], которые предполагают использование статистических показателей для обнаружения аномальной активности, таких как системные вызовы (syscalls), потребление вычислительных ресурсов оперативной памяти (RAM), графических процессоров (GPU), центральных процессоров (CPU), быстрых запросов ввода-вывода (Fast I/O request) [6], тоже недостаточно эффективны, поскольку злоумышленник при выполнении, например, сканирования открытых портов с использованием утилиты NMAP [7], может использовать параметры, задающие медленное сканирование (-T0 или -T1).

Поэтому в данной статье представлена методика и программный прототип для обнаружения аномалий в контейнерных системах на основе профилирования. Данный прототип основан на подходе поведенческого анализа, а в роли наблюдаемого объекта выступают контейнерные системы. Каждому контейнеру соответствует эталонный профиль поведения, который построен на основе последовательностей выполняемых процессов в контейнере. Обнаружение аномалий выполняется путем предсказания следующего шага на основе текущего. Если предсказанный шаг отличается от фактического следующего шага – регистрируется аномалия.

Новым элементом предлагаемого решения является создание гистограмм процессов поведения на основе трассировки системных вызовов и использования неконтролируемой гибридной модели глубокого обучения AE-LSTM, комбинирующей автокодировщик (Autoencoder, AE) и долгую краткосрочную память (Long short-term memory, LSTM). Для решения проблемы обработки большого количества данных выполняется построение гистограммы процессов, что позволят сократить количество данных для обучения модели. Гистограммы процессов определяются как вектор данных для обучения модели и последующего обнаружения. После обучения модели осуществляется обнаружение аномалий. Таким образом становится возможным обнаружить ранее неизвестные, а также аномальные последовательности выполняемых процессов в контейнерных системах.

Статья организована следующим образом. Вначале рассматриваются релевантные работы, далее излагается предлагаемый подход к обнаружению аномалий, и представляются особенности обучения и обнаружения. В следующем разделе демонстрируются

предлагаемая реализация, испытательный стенд, используемые наборы данных и результаты экспериментов. В заключении делаются основные выводы и определяются направления будущих исследований.

**1. Релевантные работы.** Большинство текущих исследований в области аномалий в контейнерных системах основаны на применении трассировки системных вызовов и различных методов глубокого машинного обучения. Они позволяют расширить существующие подходы на основе профилирования, увеличить скорость и точность обнаружения. В [8] предлагается использовать нейронную сеть LSTM для обнаружения аномального поведения путем вычисления значения разности предсказанной последующей последовательности системных вызовов на основе текущей по отношению к действительной последующей последовательности. Аномалия будет зарегистрирована при условии, если порог несовпадения будет превышать допустимый.

Работа [9] основывается на учете шаблонного потребления вычислительных ресурсов во время работы контейнерных систем. Основу предлагаемого решения составляет гибридная нейронная сеть LSTM-BLSTM, использующая комбинацию LSTM и двунаправленной сети долгой краткосрочной памяти (bidirectional LSTM, BLSTM). Предлагаемое решение работает в два этапа. Первый этап, как и в предыдущем исследовании, основывается на прогнозировании будущего значения потребления вычислительных ресурсов и производительности. Второй этап предполагает выполнение классификации ожидаемого следующего значения как нормального или аномального. По экспериментальным результатам отмечено, что предложенная модель работает лучше, чем другие рассматриваемые модели, такие как Марковские модели, рекурентные нейронные сети (Recurrent Neural Network, RNN), управляемые рекуррентные блоки (Gated Recurrent Unit, GRU), LSTM и BLSTM. Высокие результаты прогнозирования достигаются благодаря тому, что модель позволяет интегрировать не только прямые, но и обратные зависимости.

В [10] предлагается система Kubernetes Anomaly Detector (KAD). Отличительной чертой данной системы является использование различных методов машинного обучения для обнаружения различных типов аномалий. Выбор модели осуществляется автоматических из перечная доступных.

Методы глубокого машинного обучения также применяются и в задачах обнаружения вредоносного программного обеспечения. Исследование [11] направлено на обнаружение аномальных подов (pods) кластера Kubernetes, содержащих активный процесс майнингового программного обеспечения, путем анализа системных вызовов. Результаты, полученные экспериментальным путем, демонстрируют точность модели AE-LSTM - 78.9%, а точность ANN - 79,7%, соответственно. В [12] рассматривается система DockerWatch, которая выполняет обнаружение вредоносных файлов. Процесс обнаружения разбит на два этапа: первый этап предполагает быстрый анализ дизассемблированного кода и исходного двоичного файла с использованием сверточной нейронной сети (Convolutional Neural Network, CNN). Второй этап - более медленный и предполагает анализ данных последовательностей вызовов программного интерфейса приложения (application programming interface, API) с использованием гибридной нейронной сети CNN-LSTM.

Следует также выделить работы [13, 14], в которых авторы придерживаются подхода на основе трассировки системных вызовов и использовании автокодировщиков. В [13] собранная последовательность системных вызовов разделяется на несколько отдельных последовательностей. Дальнейшим шагом является построение вектора признаков путем преобразования каждой последовательности в граф. Вектор передается в модель нейронной сети автокодировщика для классификации. Классификация выполняется путем вычисления ошибки реконструкции. Если ошибка реконструкции входного вектора превышает допустимый порог, то это свидетельствует об аномалии. В [14] представлена система KubAnomaly. Как и в [13], собранная последовательность системных вызовов разделяется на небольшие временные отрезки. Для решения проблемы обработки большого количества данных принято решение собирать только четыре категории системных вызовов (файловый ввод-вывод, сетевой ввод-вывод, операции планировщики и памяти). Такое

решение позволило оптимизировать время обучения модели. Особенностью системы KubAnomaly является использование компонента "Планировщик", который управляет всеми остальными компонентами предложенной системы. Точность предложенной модели составляет 96%.

В [15] нейронная сеть AE использована для обнаружения аномальных, ранее неизвестных процессов. Предлагаемый подход основан на ошибке реконструкции, если ошибка превышает заданный порог – регистрируется аномалия.

В [16] исследуется влияние размера окна захвата системных вызовов в качестве входных данных для методов машинного обучения. В том числе исследуются и методы глубокого машинного обучения. В [17] применен метод скользящего окна с использованием гибридной нейронной сети на основе AE и LSTM.

Рассмотренные решения являются ресурсозатратными, сложными в создании и не эффективны для обнаружения атак на контейнерные системы. Следует отметить, что использование нейронных сетей, подобных LSTM и BLSTM, в совокупности с трассировкой системных вызовов без какой-либо оптимизации значительно влияет на скорость обучения модели. Таким образом, подобные решения не могут быть эффективны при внедрении в новые инфраструктуры, поскольку каждая архитектура процессора использует свой набор системных вызовов, а процессы, циркулирующие в инфраструктуре, в которую внедряется решение, могут отличаться от той, на которой проводилось обучение.

В настоящей работе предлагается решение на основе профилирования для реализации обнаружения аномального поведения с использованием гибридной нейронной сети AE-LSTM. Предложенный подход на основе профилирования позволяет значительно сократить время и затраты вычислительных ресурсов на обучение моделей нейронной сети. Также предложенное решение демонстрирует высокие показатели обнаружения при низком уровне ложных срабатываний.

**2. Предлагаемый подход.** Предлагаемое решение опирается на подход к профилированию поведения контейнерных систем и основано на трассировке системных вызовов и использовании неконтролируемой модели обучения нейронной сети AE-LSTM. Данное решение включает три основных этапа: сбор данных; нормализация; обучение и обнаружение аномалий. В рамках данной работы, под аномалией подразумевается любое отклонение от легитимной активности.

***2.1. Сбор данных.*** На первом этапе выполняется сбор данных об активности контейнерной системы. Сбор системных вызовов и их аргументов может в значительной степени отразиться на скорости обработки и нормализации данных. Поэтому, было решено игнорировать все аргументы системных вызовов.

Сбор данных осуществляется посредством Falco Security [18] с использованием модуля ядра eBPF [19], поскольку данный инструмент обладает широким набором правил, с помощью которых можно гибко настраивать процессы мониторинга контейнерных систем и наблюдать за всеми процессами, происходящими в системе. Преимуществом eBPF является то, что данная технология не требует много ресурсов, поскольку eBPF работает в пространстве ядра и данные не копируются в пространство пользователя.

В сформированном наборе данных содержатся следующие данные: метка времени, имя пользователя, идентификатор контейнера, имя контейнера, имя процесса, идентификатор процесса и системный вызов.

***2.2. Нормализация.*** Следующим этапом выступает обработка и нормализация собранной последовательности данных. Поскольку контейнер в системе не один, следует собрать последовательности системных вызовов характерных для каждого контейнера и отделить их друг от друга. Для этого необходим идентификатор процесса и контейнера.

Процесс нормализации выполняется за пять шагов: (1) модуль отвечающий за нормализацию “Normalizer” читает файл falcoDatasets.csv с последовательностями системных вызовов, построчно, опираясь на идентификатор процесса и идентификатор контейнера; (2) как только встречается идентификатор процесса, отличающийся от текущего, процесс останавливается, формируется строка с гистограммой процесса и записывается в файл характерный для контейнера (trainDatasetsContainerA.csv); имя файла формируется

на основе имени или идентификатора контейнера; (3) так продожается до тех порт пока нормализатор не встретит другой идентификатор контейнера; (4) после того, как нормализатор встретит дургой идентификатор контейнера, отличающий от текущего, все собранные последовательности буду записаны в файл, характеризующий контейнер (trainDatasetsContainerA.csv); (5) модуль нормализатора продолжает работу до тех пор, пока не встретит последнюю строку.

Таким образом, каждому контейнеру будет соответствовать CSV-файл с последовательностью выполненных процессов легитимного поведения.

Особое внимание следует уделить формированию гитограммы процесса (шаг 2 в процессе нормализации). Формирование гитограммы процесса включает два этапа.

На первом этапе выполняется оптимизация конструкций системных вызовов. Количество системных вызовов, которые поддерживаются в Falco, насчитывает 413. Для построения гистограммы необходимо создать массив размером 414 элементов, где каждый элемент этого массива по умолчанию будет равен 0, поскольку учитываются все исполняемые и неисполняемые системные вызовы. Массив размером 414 создается для того, чтобы модель нейронной сети смогла принять оптимальные параметры. Поскольку, слой, который находится между входным и выходным слоем, содержит вдвое меньше нейронов, текущий этап предполагает преобразование в числовое представление собранных системных вызовов и их подсчет.

Например, после выполнения системного вызова write (номер 4 в словаре) дважды и open (номер 1 в словаре) один раз, будет получена гистограмма (вектор), которая представлена в лист. № 1.

Листинг № 1

**Гистограмма (вектор) процесса**

```
[ 1, 0, 0, 2, 0, 0, 0, 0, 0, … ]
```

На втором этапе, как только гистограмма сформирована, она будет занесена в файл, и все последующие гистограммы выполненных процессов в контейнере также будут занесены в файл.

Таким образом, файл хранит в себе последовательность гистограмм процессов.

Данные, не используемые в процессе построения гистограмм и формирования файлов последовательностей выполненных процессов, необходимы для того, чтобы идентифицировать контейнер и дать оценку выполняемых процессов.

***2.3. Обучение и обнаружение.*** Одним из главных компонентов, реализующих предлагаемый подход, является нейронная сеть AE-LSTM. Цель данной сети заключается в обнаружении аномалий во временных рядах или последовательных данных. С этой целью сеть обучается на легитимных данных и используется для предсказания следующего шага временного ряда или последовательности данных. Предсказанное значение сравнивается со следующим фактическим значением, и вычисляется ошибка предсказания. В случае, если ошибка предсказания существенно отличается от ожидаемой, это указывает на наличие аномалии в данных. В качестве ошибки предсказания выступает пороговое значение. Если ошибка предсказания превышает заданный порог, это означает наличие аномалии в данных. Таким образом, нейронная сеть AE-LSTM обнаруживает аномалии путем анализа изменений в данных и сравнения их с ожидаемыми шаблонами.

**3. Реализация.** Сбор данных осуществляется с помощью инструмента аудита Falco Security с модулем ядра eBPF. Данные собираются с помощью правил Falco Security. Правила прописаны в файле /etc/falco/falco_rules.yaml. С помощью правил Falco можно собрать практически любую информацию о том, что происходит в системе. Но в этом нет необходимости, поскольку чем больше информации будет собираться, тем больше будет накладных расходов на вычислительные ресурсы.

После того, как набор данных сформирован, он передается в модель нейронной сети AE-LSTM для обучения.

Результатом исходного кода, представленного выше, является модель нейронной сети с использованием LSTM слоев. Сначала создается последовательная модель с несколькими слоями LSTM. Первый LSTM слой с 414 нейронами и функцией активацией ReLU принимает входные данные заданной формы (timesteps, rows) и возвращает последовательности. Следующий LSTM слой с 207 нейронами также имеет функцию активации ReLU и не возвращает последовательности. Затем добавляется слой RepeatVector для повторения входных данных timesteps еще раз. Последующие два LSTM слоя с 207 и 414 нейронами имеют функцию активацию ReLU и возвращают последовательности. Наконец, добавляется слой TimeDistributed Dense для предсказания выходных данных заданной формы. Модель компилируется с оптимизатором "adam" и функцией потерь "mse" (mean squared error). Затем модель обучается на данных X в течение 100 эпох с размером пакета 414. Эта модель предназначена для прогнозирования последовательных данных с использованием LSTM слоев и оптимизатора Adam. Представленный прототип реализован на языке Python 3 с использованием библиотек numpy, pandas, tensorflow и keras.

**4. Эксперименты**

***4.1. Испытательный стенд***. Для проведения экспериментов создан испытательный стенд, а также наборы данных легитимной, аномальной и вредоносной активности.

Испытательной стенд включает три контейнера приложения T-Pot [20]. T-Pot - это платформа, которая поддерживает более 20 "ловушек" (honeypots) и графический интерфейс для визуализации атак в режиме реального времени. Контейнеры, задействованные в эксперименте, представлены в табл. 1.

Таблица 1

**Описание контейнеров**

| Контейнер | Описание |
|---|---|
| cowrie | Cowrie - honeypot SSH / Telnet, предназначен для регистрации атак методом Brute-force |
| mailoney | Mailoney - SMTP honeypot |
| elasticsearch | Elasticsearch - honeypot, имитирующий уязвимый сервер Elasticsearch |

***4.2. Наборы данных.*** Для создания наборов данных, было использовано несколько контейнеров с имитацией легитимной, аномальной и вредоносной нагрузки. Испытательный стенд включал контейнеры с различным взаимодействием, в том числе контейнеры отправки почтовых сообщений, мониторинга системы с помощью графического интерфейса и прикладного программного обеспечения, авторизации с помощью ssh и работы в контейнере с помощью различных команд.

Информация о собранных наборах данных отображена в табл. 2.

Таблица 2

**Наборы данных**

| Группа наборов данных | Набор данных | Описание |
|---|---|---|
| A | cowrie-A | Работа в контейнере с помощью команд (sudo, ls, chmod, pwd, nano, sh, bash, touch, rm, mkdir, cd, mv), авторизация с помощью ssh |
| | mailoney-A | Отправка почтовых сообщений |
| | elasticsearch-A | Мониторинг системы, работа с графическим интерфейсом T-pot (просмотр / изменение веб – страниц) |
| B | cowrie-B | Создание запланированной задачи в контейнере, загрузка файлов с удаленного сервера |
| | mailoney-B | Сканирование портов, сканирование на наличие уязвимостей |
| | elasticsearch-B | Сканирование портов |

*Окончание табл. 2*

| | | |
|---|---|---|
| C | cowrie-C | BackConnect, Шифровальщик, Mining, Brute-force |
| | mailoney-C | DDOS |
| | elasticsearch-C | DDOS |
| D | cowrie-D | Работа в контейнере с помощью команд (sudo, ls, chmod, pwd, nano, sh, bash, touch, rm, mkdir, cd, mv), авторизация с помощью ssh, создание запланированной задачи в контейнере, загрузка файлов с удаленного сервера, BackConnect, Шифровальщик, Mining, Brute-force |
| | mailoney-D | Отправка почтовых сообщений, сканирование портов, сканирование на наличие уязвимостей, DDOS |
| | elasticsearch-D | Мониторинг системы, работа с графическим интерфейсом T-pot (просмотр / изменение веб – страниц), сканирование портов, DDOS |

Наборы данных группы “A” содержат данные легитимной активности, группы “B” – аномальной активности, “C” – вредоносной активности, а группы “D” – смешанной активности.

Каждый набор данных был сформирован за период 10 минут. В каждом наборе данных была представлена активность 10-ти пользователей. Действия пользователей были организованы в хаотичном порядке с разным интервалом между событиями, чтобы получить максимально правдоподобные наборы данных.

***4.3. Обучение моделей.*** Для обучения моделей, использовались наборы данных легитимной активности, которые представлены в табл. 2. Каждая модель соответствует профилю контейнера. Например, модель нейронной сети, скомпилированная на базе набора данных cowrie-A, соответствует профилю контейнера cowrie. Модели обучены в течении 100 эпох с размером пакета 414 и скомпилированы с оптимизатором "adam" и функцией потерь "mse".

Следует отметить, что в среднем обучение каждой модели длилось 5 минут. Обучение проводилось на испытательном стенде, который включает следующие характеристики: ОС - Cent OS 9; ОЗУ - 32 ГБ; ЦП - AMD Ryzen 5 5600X 6-Core Processor 3.70 GHz.

***4.4. Результаты.*** Обнаружение аномалий выглядит следующим образом: входной вектор последовательности гистограмм процессов подается в обученную модель нейронной сети AE-LSTM. Размер каждой последовательности составляет 10 последовательных гистограмм. На основе этих данных предсказывается следующая последовательность, если предсказанная последовательность отличается от следующей фактической последовательности, то регистрируется аномалия.

Предлагаемое решение позволяет обнаружить не только аномальные процессы, но и аномальных последовательностей процессов, что необходимо экспериментально подтвердить на разных наборах данных. Для проведения экспериментов использованы наборы данных группы “B” с аномальными данными, группы “C” с вредоносными данными и группы “D”, в которую входят легитимная, аномальная и вредоносная активность.

В результате эксперимента любая атака, аномальное или вредоносное поведение обнаруживалось как аномалия, поскольку класс атаки не определялся.

С целью экспериментального подтверждения эффективности предлагаемого решения, была проведена оценка созданных моделей на трех наборах данных B, C и D.

В каждом наборе данных содержится более 500 тысяч записей об активности на испытательном стенде. Для оценки эффективности предлагаемого решения в табл. 3 представлены результаты экспериментов на основе определения элементов матрицы ошибок (Confusion Matrix). В рамках проведенного эксперимента определялось два класса поведения: легитимное и аномалия.

Таблица 3

**Результаты проведенного эксперимента**

| Группа | Набор данных | True | | False | |
|---|---|---|---|---|---|
| | | Positive | Negative | Positive | Negative |
| B | cowrie-B | **6953** | **0** | **0** | **48** |
| | mailoney-B | **6953** | **0** | **0** | **147** |
| | elasticsearch-B | **6999** | **0** | **0** | **18** |
| C | cowrie-C | **6688** | **0** | **0** | **151** |
| | mailoney-C | **6743** | **0** | **0** | **113** |
| | elasticsearch-C | **6914** | **0** | **0** | **99** |
| D | cowrie-D | **6085** | **1519** | **202** | **291** |
| | mailoney-D | **5386** | **1805** | **80** | **213** |
| | elasticsearch-D | **2337** | **1831** | **148** | **116** |

Результаты эксперимента показывают высокую точность в обнаружении аномальных последовательностей процессов и аномальных процессов в целом, а уровень ложных срабатываний достаточно низкий.

**Заключение.** В статье представлен подход к обнаружению аномальной активности в контейнерных системах на основе профилирования действий пользователей и процессов и продемонстрирована реализация этого подхода. Подход включает в себя этапы трассировки системных вызовов, нормализации (построения гистограмм процессов), а также создания профилей легитимного поведения с помощью моделей нейронной сети AE-LSTM. Предложенный подход позволяет проводить обнаружение ранее неизвестных процессов и аномальных последовательностей. Его отличительной особенностью является построение последовательностей гистограмм процессов и использование нейронной сети AE-LSTM.

Результаты проведенных экспериментов продемонстрировали достаточно высокую точность обнаружения ранее неизвестных процессов и аномальных последовательностей процессов и низкий уровень ложных срабатываний (за счет широкого окна захвата последовательностей).

К недостаткам данного подхода можно отнести неспособность обнаруживать медленные атаки. Если атака bruteforce будет работать по таймеру (например, запрос на авторизацию будет выполняться раз в 5 секунд), подход будет не эффективным, поскольку гистограмма процесса авторизации не будет захвачена в окно последовательности. Также следует отметить, что аномальная активность, такая как, работа в нетипичное время, тоже не будет обнаружена, поскольку в рамках представленного подхода не осуществляется привязки ко времени.

Для решения задачи обнаружения аномальной активности, а именно работы в нетипичное время, планируется использовать методы, учитывающие временные ряды. Для обнаружения такой активности разрабатывается алгоритм, который отслеживает появление тех или иных гистограмм в определенный промежуток времени. Если появившаяся гистограмма не характерна для конкретного промежутка времени, будет зарегистрирована аномалия. Также в рамках будущих исследований будут протестирована эффективность обнаружения других атак (SQL-инъекции, PHP-инъекции и др.).

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Котенко И.В.* Искусственный интеллект для кибербезопасности: новая стадия противоборства в киберпространстве // Искусственный интеллект и принятие решений. – 2024. – № 1. – С. 3-19.
2. *Priedhorsky R.* Charliecloud is not affected by CVE-2024-21626 or related vulnerabilities, Los Alamos National Laboratory (LANL), Los Alamos, NM (United States), 2024, № LA-UR-24-21089.
3. *Левшун Д.С., Веснин Д.В., Котенко И.В.* Прогнозирование категорий уязвимостей в конфигурациях устройств с помощью методов искусственного интеллекта // Вопросы кибербезопасности. – 2024. – № 3 (61). – С. 33-39.

4. *Applebaum S., Gaber T., Ahmed A.* Signature-based and machine-learning-based web application firewalls: A short survey // Procedia Computer Science. – 2021. – Vol. 189. – P. 359-367.
5. *Pang G., Shen C., Cao L., Hengel A.V.D.* Deep learning for anomaly detection: A review // ACM computing surveys (CSUR). – 2021. – Vol. 54, No. 2. – P. 1-38.
6. *Ahmed M.E., Kim H., Camtepe S., Nepal S.* Peeler: Profiling kernel-level events to detect ransomware // Computer Security–ESORICS 22: 26th European Symposium on Research in Computer Security, Darmstadt, Germany, October 4–8, 2021, Proceedings, Part I 26. – Springer International Publishing, 2021. – P. 240-260.
7. *Liao S., Zhou C., Zhao Y., Zhang Z., Zhang C., Gao Y., Zhong, G.* A comprehensive detection approach of nmap: Principles, rules and experiments // 2020 international conference on cyber-enabled distributed computing and knowledge discovery (CyberC), IEEE, 2020. – P. 64-71.
8. *Snehi J., Bhandari A., Baggan V., Snehi M., Kaur H.* AIDAAS: Incident handling and remediation anomaly-based IDaaS for cloud service providers // 2021 10th International Conference on System Modeling & Advancement in Research Trends (SMART), IEEE, 2021. – P. 356-360.
9. *Gupta S., Muthiyan N., Kumar S., Nigam A., Dinesh D. A.* A supervised deep learning framework for proactive anomaly detection in cloud workloads // 2017 14th IEEE India Council International Conference (INDICON), IEEE, 2017. – P. 1-6.
10. *Kosińska J., Tobiasz M.* Detection of Cluster Anomalies With ML Techniques // IEEE Access. – 2022. – Vol. 10. – P. 110742-110753.
11. *Karn, R. R., Kudva, P., Huang, H., Suneja, S., Elfadel, I. M.* Cryptomining detection in container clouds using system calls and explainable machine learning, IEEE transactions on parallel and distributed systems, 2020, Vol. 32, №. 3, pp. 674-691.
12. *Wang Y., Wang Q., Qin X., Chen X., Xin B., Yang R.* DockerWatch: a two-phase hybrid detection of malware using various static features in container cloud // Soft Computing. – 2023. – Vol. 27, No. 2. – P. 1015-1031.
13. *El Khairi A., Caselli M., Knierim C., Peter A., Continella A.* Contextualizing system calls in containers for anomaly-based intrusion detection // Proceedings of the 2022 on Cloud Computing Security Workshop. – 2022. – P. 9-21.
14. *Tien C.W., Huang T.Y., Tien C.W., Huang T.C., Kuo S.Y.* KubAnomaly: Anomaly detection for the Docker orchestration platform with neural network approaches // Engineering reports. – 2019. – Vol. 1, No. 5. P. e12080.
15. *Kotenko I., Melnik M., Abramenko G.* Anomaly detection in container systems: using normal process histograms and an autoencoder // 2024 IEEE 25th International Conference of Young Professionals in Electron Devices and Materials (EDM 2024), IEEE, 2024. – P. 1930-1934.
16. *Castanhel G.R., Heinrich T., Ceschin F., Maziero C.* Taking a peek: An evaluation of anomaly detection using system calls for containers // 2021 IEEE Symposium on Computers and Communications (ISCC), IEEE, 2021. – P. 1-6.
17. *Cui P., Umphress D.* Towards unsupervised introspection of containerized application // Proceedings of the 2020 10th International Conference on Communication and Network Security. – 2020. – P. 42-51.
18. *Deri L., Sabella S., Mainardi S., Degano P., Zunino, R.* Combining System Visibility and Security Using eBPF // ITASEC. – 2019. – Vol. 2315. – P. 1-12.
19. *Kotenko I., Saenko I., Chechulin A., Vitkova L., Kolomeec M., Zelichenok I., Melnik M., Makrushin D., Petrevich N.* Detection of Anomalies and Attacks in Container Systems: An Integrated Approach Based on Black and White Lists // International Conference on Intelligent Information Technologies for Industry, Cham: Springer International Publishing, 2022. – P. 107-117.
20. *Subhash P., Qayyum M., Likhitha Varsha C., Mehernadh K., Sruthi J., Nithin A.* A Security Framework for the Detection of Targeted Attacks Using Honeypot // International Conference on Computer & Communication Technologies, Singapore: Springer Nature Singapore, 2023. – P. 183-192.

REFERENCES

1. *Kotenko I.V.* Iskusstvennyy intellekt dlya kiberbezopasnosti: novaya stadiya protivoborstva v kiberprostranstve [Artificial Intelligence for Cybersecurity: A New Stage of Confrontation in Cyberspace], *Iskusstvennyy intellekt i prinyatie resheniy* [Artificial Intelligence and Decision Making], 2024, No. 1, pp. 3-19.
2. *Priedhorsky R.* Charliecloud is not affected by CVE-2024-21626 or related vulnerabilities, Los Alamos National Laboratory (LANL), Los Alamos, NM (United States), 2024, No. LA-UR-24-21089.
3. *Levshun D.S., Vesnin D.V., Kotenko I.V.* Prognozirovanie kategoriy uyazvimostey v konfiguratsiyakh ustroystv s pomoshch'yu metodov iskusstvennogo intellekta [Forecasting vulnerability categories in device configurations using artificial intelligence methods], *Voprosy kiberbezopasnosti* [Cybersecurity Issues], 2024, No. 3 (61), pp. 33-39.

4. *Applebaum S., Gaber T., Ahmed A.* Signature-based and machine-learning-based web application firewalls: A short survey, *Procedia Computer Science*, 2021, Vol. 189, pp. 359-367.
5. *Pang G., Shen C., Cao L., Hengel A.V D.* Deep learning for anomaly detection: A review, *ACM computing surveys (CSUR),* 2021, Vol. 54, No. 2, pp. 1-38.
6. *Ahmed M.E., Kim H., Camtepe S., Nepal S.* Peeler: Profiling kernel-level events to detect ransomware, *Computer Security–ESORICS 22: 26th European Symposium on Research in Computer Security, Darmstadt, Germany, October 4–8, 2021, Proceedings, Part I 26, Springer International Publishing, 2021*, pp. 240-260.
7. *Liao S., Zhou C., Zhao Y., Zhang Z., Zhang C., Gao Y., Zhong, G.* A comprehensive detection approach of nmap: Principles, rules and experiments, *2020 international conference on cyber-enabled distributed computing and knowledge discovery (CyberC), IEEE, 2020*, pp. 64-71.
8. *Snehi J., Bhandari A., Baggan V., Snehi M., Kaur H.* AIDAAS: Incident handling and remediation anomaly-based IDaaS for cloud service providers, *2021 10th International Conference on System Modeling & Advancement in Research Trends (SMART), IEEE, 2021*, pp. 356-360.
9. *Gupta S., Muthiyan N., Kumar S., Nigam A., Dinesh D.A.* A supervised deep learning framework for proactive anomaly detection in cloud workloads, *2017 14th IEEE India Council International Conference (INDICON), IEEE, 2017*, pp. 1-6.
10. *Kosińska J., Tobiasz M.* Detection of Cluster Anomalies With ML Techniques, *IEEE Access*, 2022, Vol. 10, pp. 110742-110753.
11. *Karn R.R., Kudva, P., Huang H., Suneja S., Elfadel I.M.* Cryptomining detection in container clouds using system calls and explainable machine learning, *IEEE transactions on parallel and distributed systems*, 2020, Vol. 32, No. 3, pp. 674-691.
12. *Wang Y., Wang Q., Qin X., Chen X., Xin B., Yang R.* DockerWatch: a two-phase hybrid detection of malware using various static features in container cloud, *Soft Computing*, 2023, Vol. 27, No. 2, pp. 1015-1031.
13. *El Khairi A., Caselli M., Knierim C., Peter A., Continella A.* Contextualizing system calls in containers for anomaly-based intrusion detection, *Proceedings of the 2022 on Cloud Computing Security Workshop*, 2022, pp. 9-21.
14. *Tien C.W., Huang T.Y., Tien C.W., Huang T.C., Kuo S.Y.* KubAnomaly: Anomaly detection for the Docker orchestration platform with neural network approaches, *Engineering reports*, 2019, Vol. 1, No. 5, pp. e12080.
15. *Kotenko I., Melnik M., Abramenko G.* Anomaly detection in container systems: using normal process histograms and an autoencoder, *2024 IEEE 25th International Conference of Young Professionals in Electron Devices and Materials (EDM 2024), IEEE. 2024*. pp. 1930-1934.
16. *Castanhel G. R., Heinrich T., Ceschin F., Maziero C.* Taking a peek: An evaluation of anomaly detection using system calls for containers, *2021 IEEE Symposium on Computers and Communications (ISCC), IEEE, 2021*, pp. 1-6.
17. *Cui P., Umphress D.* Towards unsupervised introspection of containerized application, *Proceedings of the 2020 10th International Conference on Communication and Network Security*, 2020, pp. 42-51.
18. *Deri L., Sabella S., Mainardi S., Degano P., Zunino, R.* Combining System Visibility and Security Using eBPF, *ITASEC*, 2019, Vol. 2315, pp. 1-12.
19. *Kotenko I., Saenko I., Chechulin A., Vitkova L., Kolomeec M., Zelichenok I., Melnik M., Makrushin D., Petrevich N.* Detection of Anomalies and Attacks in Container Systems: An Integrated Approach Based on Black and White Lists, *International Conference on Intelligent Information Technologies for Industry, Cham: Springer International Publishing*, 2022, pp. 107-117.
20. *Subhash P., Qayyum M., Likhitha Varsha C., Mehernadh K., Sruthi J., Nithin A.* A Security Framework for the Detection of Targeted Attacks Using Honeypot, *International Conference on Computer & Communication Technologies, Singapore: Springer Nature Singapore*, 2023. pp. 183-192.

Статью рекомендовал к опубликованию д.т.н., профессор Ю.А. Кравченко.

**Котенко Игорь Витальевич** – Санкт-Петербургский федеральный исследовательский центр Российской академии наук; e-mail: ivkote@comsec.spb.ru; г. Санкт-Петербург; Россия; лаборатория проблем компьютерной безопасности; д.т.н.; профессор.

**Мельник Максим Владимирович** – e-mail: mkmxvh@gmail.com; лаборатория проблем компьютерной безопасности; аспирант.

**Kotenko Igor Vitalievich** – Saint Petersburg Federal Research Center of the Russian Academy of Sciences; e-mail: ivkote@comsec.spb.ru; Saint Petersburg; Russia; Laboratory of computer security problems; dr. of eng. sc.; professor.

**Melnik Maxim Vladimirovich** – e-mail: mkmxvh@gmail.com; Laboratory of computer security problems; postgraduate student.

УДК 004.415.2.031.43 DOI 10.18522/2311-3103-2024-5-48-58

**И.И. Левин, Е.А. Дудников**

# СТРУКТУРНАЯ МОДИФИКАЦИЯ МЕТОДА ХАФФМАНА ДЛЯ СЖАТИЯ ПЛОТНЫХ ПОТОКОВ ДАННЫХ БЕЗ ПОТЕРЬ НА РВС

*Современные запросы общества требуют решения целого ряда вычислительно трудоемких задач в режиме реального времени. Для подобных решений необходимы огромные вычислительные мощности, широкополосные высокоскоростные каналы передачи данных и внушительные объемы памяти. Обеспечить подобные запросы можно за счет разработки и внедрения новых технологий, наращивания технической инфраструктуры, что потребует значительных финансовых и временных затрат. Облегчить подобный переход, используя существующую техническую базу, можно за счет использования алгоритмов сжатия данных в реальном времени. Средства сжатия данных в темпе поступления могут повысить скорость вычислений, передачи данных, снизить занимаемый объем при хранении, используя имеющуюся инфраструктуру. Современные технические платформы на базе CPU не способны обеспечить потоковую обработку данных в темпе их поступления, реальная производительность подобных систем не превышает 10 % от пиковой. Новой платформой для систем сжатия данных без потерь в темпе поступления могут стать реконфигурируемые вычислительные системы (РВС) на базе программируемых логических интегральных схем (ПЛИС). Однако для эффективной работы подобных систем требуется разработка новых методов с применением структурных вычислений, позволяющих полностью раскрыть потенциал ресурса ПЛИС. В данной работе представляется реализация на РВС модификации динамического алгоритма кодирования Хаффмана, которая позволяет создавать префиксные коды оптимальной длины и обрабатывать плотные потоки данных в темпе поступления с пропускной способностью не менее 128 Гбит\с. Производительность разработанной модификации в 5 раз превосходит наилучшую известную комплементарную реализацию на базе FPGA на один вычислительный конвейер.*

*РВС; динамическое кодирование Хаффмана; обработка данных в реальном времени.*

**I.I. Levin, E.A. Dudnikov**

# STRUCTURAL MODIFICATION OF THE HUFFMAN METHOD FOR COMPRESSION OF DENSE DATA STREAMS WITHOUT LOSS ON A RCS

*Modern society demands require solving a whole range of computationally intensive tasks in real time. Such solutions require enormous computing power, broadband high-speed data transmission channels and impressive memory capacities. Such demands can be met by developing and implementing new technologies, expanding the technical infrastructure, which will require significant financial and time costs. Such a transition can be facilitated using the existing technical base by using real-time data compression algorithms. Data compression tools at the rate of receipt can increase the speed of calculations, data transfer, and reduce the occupied space during storage, using the existing infrastructure. Modern CPU-based technical platforms are not capable of providing streaming data processing at the rate of their receipt; the actual performance of such systems does not exceed 10% of the peak. Reconfigurable computing systems (RCS) based on programmable logic integrated circuits (FPGAs) can become a new platform for lossless data compression systems at the rate of receipt. However, for the efficient operation of such systems, it is necessary to develop new methods using structural calculations that allow the full potential of the FPGA resource to be unleashed. This paper presents the implementation of a modification of the dynamic Huffman coding algorithm on the RCS, which allows creating prefix codes of optimal length and processing dense data flows at the rate of receipt with a throughput of at least 128 Gbit/s. The performance of the developed modification is 5 times higher than the best known complementary implementation based on FPGA per computing pipeline.*

*RCS; dynamic Huffman coding; real-time data processing.*

**Введение.** Современные требования к качеству решения вычислительно трудоёмких научно-технических задач неуклонно ведёт к увеличению объёмов производимой и обрабатываемой информации. Так по оценкам специалистов человечеством производится порядка 320 миллионов терабайт данных ежедневно, а общий объём произведённой информации к 2025 году превысит 175 зеттабайт [1]. Согласно этим отчётам объем репли-

цированной информации в девять раз выше объема оригинальной произведенной информации, и подобное смещение будет только нарастать. Большая часть скопированной и использованной информации носит временный характер и зачастую служит для одноразового использования. Растёт объём метаданных и данных различных сервисов о конкретных пользователях, который будет значительно превышать объём данных, созданный этими пользователями.

Немаловажную роль в подобном бурном росте сыграла разразившаяся в 2019 году пандемия, потребовавшая перехода множества организаций, связанных с производством и образованием, на удалённые формы взаимодействия. Основными отраслями, генерирующими информацию, по-прежнему остаются промышленность, здравоохранение, финансовый сектор, развлечения и медиабизнесс [2]. Благодаря снижению стоимости пользования облачными хранилищами происходит замена физических серверов. Уже сегодня около 94% крупных компаний используют облачные хранилища, а в 2025 году туда может переместиться до 80% всей рабочей нагрузки в мире [3]. Всё это ведёт к необходимости хранения, передачи и обработки огромного объема данных в том числе и в режиме «реального времени». К отраслям, требующим обработки данных в реальном времени, можно отнести военную, космическую области, задачи ядерной физики, радиолокации, геолокации, предиктивной аналитики, промышленные автоматические системы управления производством, энергетику, трансляции потоков аудиовизуальной информации от источника к конечному пользователю (мультимедиа, телекоммуникации) [4].

Непременными требованиями к задачам реального времени являются не только корректность вычислений, но и временные рамки. Латентность системы может быть различной, но она должна быть гарантированной и предсказуемой - нарушение сроков исполнения задачи расценивается как возможный отказ системы. Принадлежность системы к классу задач реального времени никак не связана с ее быстродействием: время обработки может быть различным, требования сроков исполнения задачи определяется техническим заданием и логикой функционирования системы [5].

Вычислительно трудоёмкие задачи реального времени и растущие объёмы данных требуют увеличения мощности вычислительных систем, объемов общей и оперативной памяти, увеличения пропускной способности каналов передачи данных. Для этого необходимо перестроить значительную часть существующей инфраструктуры. Это влечёт за собой существенные финансовые и временные затраты на разработку и внедрение новых технологий обработки, передачи и хранения информации. Облегчить подобный переход могут новые высокоскоростные системы сжатия данных, позволяющие более эффективно использовать уже существующую техническую базу. Уже сейчас крупные компании внедряют новые алгоритмы сжатия данных для облачных хранилищ [6]. Одним из основных требований к подобным системам является возможность восстановления исходных данных без потерь.

Для решения трудоёмких задач выделяют основные типы вычислительных устройств: центральные процессоры (CPU), графические процессоры (GPU) специализированные устройства (ASIC) и программируемые вентильные матрицы (FPGA). Эти устройства имеют между собой явные различия. Большинство задач так или иначе можно выполнять с помощью CPU, однако другие типы процессоров позволяют решать многие вычислительно сложные задачи быстрее с меньшими энергозатратами.

По причине широкого распространения большая часть задач по сжатию данных попрежнему выполняется системами на базе CPU. Однако подобные системы не способны обеспечить требуемый уровень производительности. Исследования показывают, что при сжатии данных алгоритмами семейства DEFLATE производительность ядра составляет от нескольких десятков до нескольких сотен Мбит/с. При использовании специализированных многоядерных систем независимо от типа многоядерной машины и числа ядер скорость обработки потока составляет от 512 Мбит/с до 3,4 Гбит/с в зависимости от коэффициента сжатия [7]. Наращивание аппаратного ресурса не приводит к линейному

росту производительности, что связано с проблемами распараллеливания процессов, задействованности ядер и межпроцессорными связями. Причина кроется в самой структуре универсальных процессоров, которые не предназначены для высокопроизводительных систем с конвейерной организацией вычислений.

Эффективность GPU при переносе на них классических методов сжатия без потерь показывает сопоставимые с CPU результаты 2,5–7,3 Гбит/с. Применение GPU позволяет повысить производительность за счёт вычислительных мощностей, но это требует разработки и адаптации существующих методов сжатия. Использование специализированных адаптированных методов на основе существующих, предварительная обработка данных (например, преобразование Барроуза-Уилера) [8] позволяет повысить производительность GPU, но подобный рост также может приводить к кратному снижению степени сжатия и росту объёма вычислений.

Технической основой для устройств по сжатию данных могут быть специализированные устройства на базе заказных кристаллов или реконфигурируемые вычислительные системы (РВС) на базе FPGA. Производительность коммерческих реализаций методов сжатия данных без потерь на базе ASIC колеблется в пределах 20 Гбит/с на кристалл [9]. Разработка подобных систем является долгим и затратным предприятием. Более того, при внесении изменений в алгоритм, обнаружении ошибки, модернизации потребуется повторная разработка и перевыпуск всей серии устройств. Производительность комплементарных систем на базе CPU и FPGA не превышают 120,4 Гбит/с на один кристалл. Наилучшей реальной производительности одного вычислительного конвейера в 20 Гбит/с добились специалисты CDSC и Xilinx [10]. Добиться подобной производительности разработчикам удалось за счет применения статических энтропийных методов. Подобные решения эффективны лишь для заранее известных типов данных, близких по содержанию к эталону. При обработке разнородных данных без использования подготовленных статистических таблиц такое решение приведет к значительному снижению степени сжатия входного сообщения.

Достичь оптимального уровня сжатия для алгоритмов кодирования данных в темпе их поступления на РВС с сохранением высокой пропускной способности можно с применением принципов структурных вычислений. Однако традиционные методы архивирования не учитывают специфику работы РВС, поэтому для эффективной реализации алгоритмов сжатия данных необходимо разработать новые методы управления потоками данных и организации вычислений. Применение подобных методов позволит превзойти производительность уже существующих систем сжатия данных без потерь на РВС в темпе поступления и достичь производительности 128 Гбит/с и более на один конвейер. Предполагается, что разработка и внедрение новых методов позволит сократить задействованный аппаратный ресурс и сократить латентность вычислений по сравнению с известными методами. На основе вышеизложенного можно сделать заключение, что разработка новых методов эффективного сжатия высокоскоростных потоков данных в темпе их поступления на РВС является актуальной задачей.

**Постановка задачи.** Современные системы сжатия данных без потерь (Deflate, Snappy, Gipfeli, Zopfli, Brotli и т.д.) являются комплексными, включают несколько различных алгоритмов преобразования и кодирования. Важным элементом подобных систем являются энтропийные алгоритмы, позволяющие генерировать префиксные коды оптимальной длины. Энтропийное кодирование основано на частоте повторения символов в сообщении и направленно на устранение избыточности и уменьшение занимаемого сообщением объема.

Одним из наиболее популярных энтропийных методов, широко используемым во многих архиваторах и протоколах данных, является разработанный в 1952 году метод Хаффмана [11]. Идея метода заключается в том, что символу алфавита сообщения $s_i \in \{s_1, s_2, \dots, s_m\}$ с фиксированной длинной и соответствующей частотой появления $w_i \in \{w_1, w_2, \dots, w_m\}$, $w_1 \le w_2 \le \dots \le w_m$ присваивается новый код переменной длины $l_i$, зависящей от $w_i$ так, что чем выше частота появления, тем короче новый код символа. Тогда минимизированную длину нового сообщения можно представить как выра-

жение $\sum_{i=1}^{m} w_i l_i$. являющимся производным от уравнения средней длины кода Шеннона [12]. Полученное сообщение может быть однозначно декодировано благодаря префиксности кода.

Существующие алгоритмы кодирования Хаффмана можно разделить на статические, динамические и адаптивные. Динамический метод требует двух вычиток передаваемого сообщения из памяти: первый раз для сбора статистики и генерации кодов, второй раз для кодирования. Этот алгоритм обеспечивает оптимальную степень сжатия кодируемого сообщения. В случае, когда можно заранее спрогнозировать распределение вероятностей появления символов в сообщении, обосновано применение статического метода Хаффмана с применением заранее рассчитанных таблиц кодов. Такой подход не позволяет создавать коды оптимальной длины, что снижает степень сжатия. Степень сжатия снижается еще сильнее, если обрабатываются разнородные или неизвестные заранее данные. Для некоторых сфер, требующих обработки данных в темпе их поступления, таких, например, как телекоммуникационные системы, статический метод Хаффмана может оказаться малоэффективным. Для подобных задач применяют адаптивные варианты алгоритма Хаффмана. Адаптивный метод позволяет решать задачу за одно прочтение без использования таблиц статистики. Однако по сравнению с динамическим, адаптивный алгоритм требует выполнения большего числа операций, и обладает более низким коэффициентом сжатия данных. Исходя из соотношения степени сжатия к сложности вычислений, можно сделать вывод, что при обработке больших массивов разнородных данных динамический алгоритм Хаффмана является более эффективным по сравнению с адаптивным или статическим алгоритмом. Однако необходимость двойного чтения сообщения из памяти не позволяет кодировать плотные потоки данных на традиционных вычислительных системах в режиме реального времени.

Известен ряд модификаций динамического алгоритма Хаффмана, предлагающих построение префиксных кодов с минимальной избыточностью (Minimum-Redundancy Codes), направленных на снижение вычислительной сложности и уменьшение аппаратных затрат. Например, в работах [13, 14] предлагается параллельно заполнять многомерный массив дерева Хаффмана и рассчитывать длины кодов с последующей генерацией кодов за линейное время. В работе [15] предложено построение дерева Хаффмана без создания и хранения таблицы массива данных, описывающих дерево и прохода по этой таблице при генерации новых кодов, а путем расчета длин кодов и последующей генерацией кодов. В зависимости от используемого метода полученные коды могут отличаться, но уровень энтропии новых кодов и минимальная избыточность полученных кодов сохраняется для каждого символа алфавита согласно неравенству Крафта $K = \sum_{i=1}^{m} 2^{-l_i} \leq 1$, где $l_i$ это число бит в коде символа (длина кода) [16]. Тем не менее, подобные модификации не избавляют от недостатков динамического алгоритма Хаффмана, а именно необходимости двойного чтения кодируемого сообщения.

Структурная реализация алгоритма Хаффмана на РВС позволит избавиться от подобных проблем. В частности, блок буферизации и синхронизации входного сообщения позволит реализовать динамический алгоритм Хаффмана на РВС, выполняя лишь одну загрузку сообщения из внешнего источника. Оценив максимальную пропускную способность известных реализаций на базе FPGA, были выдвинуты технические требования для разрабатываемой модификации. Структура должна обрабатывать плотные потоки данных на скорости 128 Гбит/с. Для соответствия подобным требованиям при ширине шины данных 512 бит рабочая частота системы должна быть не ниже 250 МГц. При подобном темпе поступления и размере пакета 64 Кбайт (максимальный размер IP-датаграммы) для обработки сообщения потребуется 1024 такта, что и будет являться интенсивностью (скважностью) системы. Соответственно, статистика появления символов в сообщении для каждого блока должна формироваться один раз в 1024 такта. Использование известных последовательных алгоритмов не позволит должным образом распараллелить задачу. Кроме того, задача кодирования методом Хаффмана относится к классу задач с переменной интенсивностью потоков данных и последовательная обработка или вызывает

нерегулярные задержки, или увеличивает общую латентность решения до максимальной. При этом система буферизации и синхронизации, как для классического динамического алгоритма Хаффмана, так и его модификаций может занять значительный аппаратный ресурс, который может превосходить затраты вычислительной части алгоритма. Глубина подобной системы непосредственно зависит от латентности блока расчета и построения таблицы новых кодов, поэтому целесообразно использовать модификацию, обеспечивающую наибольшую скорость исполнения.

Процесс получения новых кодов можно представить как заполнение двумерной матрицы $A_{(n,m)}$, где m – это количество строк, соответствующее числу символов алфавита, имеющих вес, а n - наибольшая ширина машинного слова (длина кода). Было рассчитано, что для блока сообщения в 65536 символов наибольшая длина кода $n_{max}$ =22 при вырожденном дереве, отображающем последовательность Фибоначчи:

$$F_n = F_{n-1} + F_{n-2},$$

где $F_0, F_1, F_3 = 1$ и $F_n \leq 65536$.

Для сокращения вычислительной нагрузки некоторые алгоритмы ограничивают длину кода. Согласно уравнению средняя длина кода $K = \log_2 w_{max}$, где $w_{max}$ – максимально допустимый вес символа. При $w_{max} = 65636$ средняя длина кода составит $K = 16$, что совпадает с регламентами некоторых технических протоколов [17].

**Описание модифицированного метода Хаффмана с разработанными блоками построения дерева и расчета новых кодов.** Классический метод Хаффмана предполагает заполнение таблицы узлов бинарного дерева с последующим вычислением кодов путём либо прямого прохода от вершины к корню, либо рекурсивного прохода со ссылками на предыдущие узлы. Сложность динамического алгоритма Хаффмана в значительной степени обуславливается числом обращений к многомерному массиву предварительно построенного дерева, хранение которого требует значительных аппаратных затрат. Максимальное число обращений к узлам таблицы кодов определяет сложность алгоритма и формирует минимальную латентность процесса генерации кодов $S_{lat}$. При моделировании классического метода реальная латентность блоков построения дерева Хаффмана и создания новых кодов составила 5734 такта. Для эффективной работы исходный граф был векторизирован [18] с целью соответствия техническим требованиям задачи кодировки данных в темпе их поступления. Блоки сбора статистики и сортировки обрабатывают входное сообщение за 1024 такта и формируют скважность поступления пакетов $S_p$. Для успешной работы необходимо, чтобы скважность работы структуры $G_{code}$ отвечала условию $S_{code} \leq S_p$. В полученном векторизованном графе скважность работы блока будет равна его латентности $S_{code} = S_{lat}$. Если $S_{code} > S_p$, то уменьшить скважность системы построения кодов можно за счёт распараллеливания на верхнем иерархическом уровне путем параллельного подключения нескольких макроконвейеров [19] $G_{code}$. Реализация классического метода Хаффмана потребует параллельной установки пяти макроконвейеров (рис. 1).

Было рассчитано, что распараллеливание на нижнем (итерационном) иерархическом уровне занимает меньший ресурс по сравнению с распараллеливанием по верхнему уровню с сохранением функциональности. При сохранении условия $S_{code} \leq S_p$ суммарный аппаратный ресурс будет сокращён до минимума. Анализ метода, описанного работе [20], позволил определить, что распараллеливание по каналам доступа к памяти позволяет уменьшать латентность вычислительной структуры вместе с увеличением интенсивности обработки информации. Было выполнено распараллеливание вычислительной структуры классического алгоритма построения дерева Хаффмана по каналам доступа к памяти родителей, направленное на уменьшение её латентности, что позволило существенно уменьшить аппаратные затраты на реализацию подсистемы синхронизации.

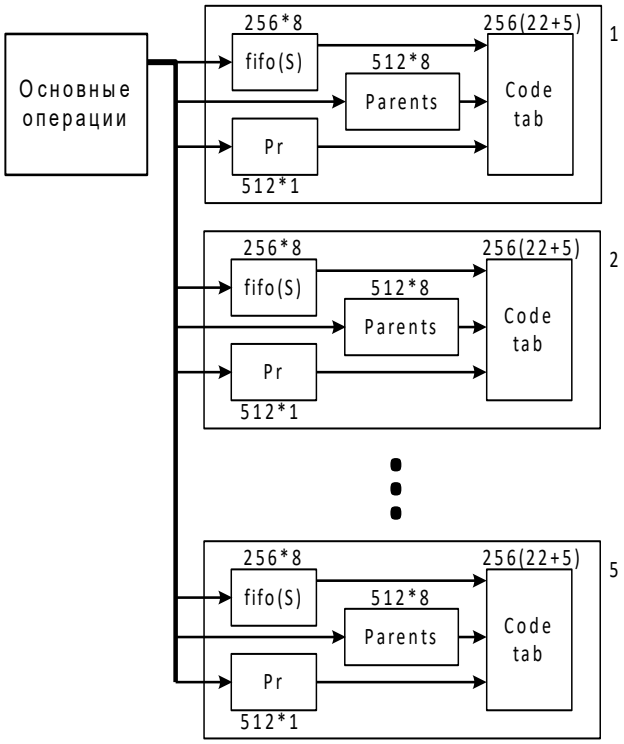



*Рис. 1. Макроконвейерная реализация вычислительной структуры при реализации классического алгоритма построения дерева Хаффмана на РВС*

Моделирование модификации классического метода Хаффмана показало, что применение параллельного доступа к памяти позволило снизить латентность вычислительной структуры $S_{code}$ до 1792 тактов. Поскольку $S_{code}$ превышает скважность поступающих данных, то для организации обработки входящего сообщения в темпе поступления требуется установка двух параллельных вычислительных блоков на макроуровне. Дальнейшее снижение латентности вычислительного блока потребовало разработки новой модификации динамического алгоритма.

В новой разработанной модификации Хаффмана, информационный граф которой представлен на рис. 2, процессы построения дерева и вычисления кодов выполняются параллельно. Подобный подход позволяет получать коды без предварительного построения дерева, заполнения и хранения таблицы узлов с ссылками на родителей, а также многократного прохода по некоторым ветвям при генерации кода для каждого символа алфавита.

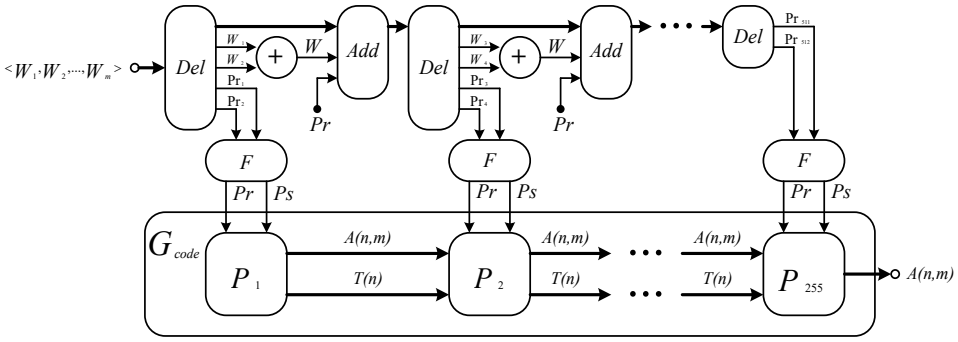



*Рис. 2. Информационный граф разработанной модификации метода Хаффмана с распараллеливанием по каналам доступа к памяти*

В основе новой модификации лежит понимание того, что дерево Хаффмана можно представить в виде набора уровней, а информации о числе уровней дерева и их наполнении достаточно для построения новых кодов и расчета их длин. Основу дерева представляют элементы, родителем которых является корневая вершина дерева. Верхний уровень дерева составляют листья с наименьшим весом. Любое подобное дерево может мыть описано совокупностью элементов на каждом из уровней. Важным изменением при распараллеливании по числу каналов доступа к памяти является преобразование информационного графа $G_{code} = \bigcup_{j=1}^{k} P_j$, где $k = (m-1)$, и подграфа $P$ для обработки не одного, а пары символов. Операции $Del$ последовательно удаляют из массива весов $< W_1, W_2, \dots, W_m >$ два элемента. Веса удалённых элементов складываются, сумма получает признак и как новый элемент помещается в массив операцией $Add$. Признак каждого элемента поступает в подграф $P$ и обрабатывается подграфами $L$, располагающиеся внутри графа $P$ в виде равномерно расширяющегося к низу дерева.

Это нововведение позволяет одновременно писать признаки данных по двум выходящим веткам на каждом уровне дерева, что сокращает общее время построения кодов. В зависимости от того, чем являются удаляемые элементы, в блок построения дерева передается признак $P(j)$:

- при $P(j) = 2$ – оба удаляемых элемента являются символами;
- при $P(j) = 1$ – один элемент является символом, другой – узлом;
- при $P(j) = 0$ – оба удаляемых элемента являются узлами.

В блоке построения дерева, показанном на рис. 3, представлены новые разработанные блоки, которые рассчитывают общее число потомков-символов и потомков-узлов после каждого родителя, расположенных уровнем выше. Эти суммы признаков символов $C(j)$ и узлов $U(j)$, полученные на каждой итерации алгоритма, а также исходные коды символов $W_1 \div W_m$, расположенные в порядке увеличения их весов накапливаются в памяти $BUF1$.

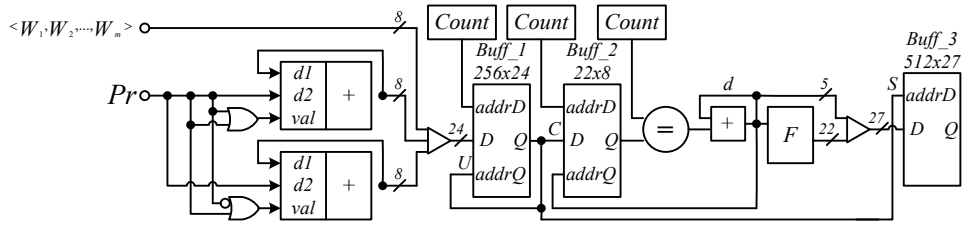



*Рис. 3. Реализация на РВС вычислительной структуры блока построения дерева Хаффмана*

После подсчета всех признаков элементов дерева определяется число символов, расположенных ниже каждого из родителей $A(i)$, где $i \le 22$ – предельное количество уровней в дереве Хаффмана для сообщения 64Кбайт. Это значение накапливается в памяти $BUF2$. Данный подход организован с применением двух очередей в памяти FIFO. Первая – это очередь признаков, указывающая номера строк для записи, которую также помещают в очередь и достают оттуда при получении последующих признаков со значением «0» для следующего столбца. Вторая очередь представляет собой производные от позиции Ps данные, из которых формируется новый код. Поскольку дерево Хаффмана обладает строгой вертикальной иерархией от корневого узла к листьям, то на конкретных этапах $j$ сумма признаков символов $C(j)$ определяет общее число символов, расположенных ниже родителей на каждом из уровней. Количество узлов $U(j)$, расположенных ниже родителей на каждом из уровней, определяет операции, которые необходимо провести на каждом этапе $j$. На основе полученных данных $A(i)$ в блоке $F$ рассчитывается новый код $e$ с соответствующей длиной кода $d$ и записывается в таблицу перекодировки $BUF3$. позиция символа в отсортированном массиве данных в зависимости от его веса. Оценка

вычислительной трудоёмкости при моделировании разработанного алгоритма показала, что латентность полученной модификации $S_{code} = 1024$ и не превышает скважность поступления данных $S_p$.

В процессе анализа были синтезированы и промоделированы несколько модификаций динамического алгоритма Хаффмана: «классический» алгоритм, модификация Хашемиана, модификация Алексеева, соответствующая по вычислительной сложности модификации Моффата, а также новый разработанный алгоритм. Все модификации, реализованные на ПЛИС XC7K355T фирмы Xilinx серии Kintex 7, обеспечили кодирование плотного потока входящих данных со скоростью 128 Гбит/с и частотой работы кристалла 250 МГц. Сравнение характеристик различных модификаций представлены в табл. 1.
P(lut), P(bram) – характеристики, показывающие долю аппаратного ресурса, занимаемую вычислительной структурой рассмотренных алгоритмов построения дерева Хаффмана, относительно ресурса, занимаемого классическим алгоритмом.

В табл. 2 представлены аппаратные затраты, используемые на реализацию рассмотренных алгоритмов динамического алгоритма кодирования Хаффмана с учётом блоков подсчёта вероятностей и сортировки, по таким параметрам как количество используемых логических таблиц (LUT), количество триггеров (Flip-Flops) и количество блоков встроенной памяти ПЛИС объемом 36Кб (BRAM).

Таблица 1

**Сравнение различных реализаций блоков построения дерева Хаффмана и генерации кодов**

| параметр | latency | LUT | BRAM, huff | BRAM, sync | BRAM, total | P(lut) | P(bram) |
|---|---|---|---|---|---|---|---|
| Классический | 5734 | 1051 | 20 | 171 | 191 | 1 | 1 |
| Алексеев\Моффат | 3968 | 3476 | 2 | 114 | 116 | 3.31 | 0.61 |
| Модифицир. классический | 1792 | 1124 | 18 | 57 | 75 | 1.07 | 0.4 |
| Хашемиан | 1280 | 199 | 3 | 44 | 47 | 0.2 | 0.25 |
| Разработанный | 1024 | 177 | 3 | 29 | 32 | 0.17 | 0.17 |

Таблица 2

**Латентность и аппаратные затраты на реализацию рассматриваемых модификаций динамического алгоритма Хаффмана**

| параметр | latency | LUT | BRAM,36K | Flip-Flops |
|---|---|---|---|---|
| Классический | 5734 | 50859 | 170,5 | 81375 |
| Алексеев\Моффат | 3968 | 53284 | 133 | 87919 |
| Хашемиан | 1280 | 50007 | 98,5 | 80011 |
| Разработанный | 1024 | 49985 | 90,5 | 74977 |

Реализация любого из рассмотренных алгоритмов позволяет достигать необходимого уровня производительности вычислительной структуры для решения поставленной задачи, однако анализ полученных результатов показал, что вычислительная структура разработанной модификации имеет наименьшую латентность (latency) и требует меньше аппаратных затрат по сравнению с другими модификациями. При масштабировании вычислительной структуры производительность РВС будет увеличиваться линейно, что позволит при необходимости выполнять кодирование потоков данных на большей скорости.

**Заключение.** Представленная в данной статье разработанная модификация построения дерева Хаффмана и расчёта новых кодов позволяет эффективно реализовать динамический алгоритм кодирования Хаффмана на РВС. Согласно парадигме структурных вычислений данная реализация позволяет убрать нелинейность структуры за счёт

разбиения на итерационные уровни и добиться высокого уровня распараллеливания задачи. Снижение вычислительной сложности и латентности вычислительной структуры позволило при реализации на ПЛИС повысить удельную производительность с сохранением заданной скорости поступления потока данных. Более того, новые коды рассчитываются динамически, что позволяет получать новые коды оптимальной длины для текущего сообщения и обеспечивает максимальную степень сжатия. За счет системы буферизации и синхронизации достаточно вычитывать входное сообщение лишь один раз, что снижает межпроцессорные обмены. Предложенное решение позволяет обрабатывать высокоскоростные плотные потоки данных в темпе поступления.

Анализ аппаратных затрат и вычислительной сложности показал, что снижение вычислительной сложности и латентности в разработанной модификации позволило уменьшить занимаемый вычислительным блоком аппаратный ресурс, сократить объем используемой внутренней памяти ПЛИС на систему буферизации и отказаться от распараллеливания на макроуровне.

Было выполнено функциональное моделирование предложенного решения. Производительность новой модификации в 5 раз превосходит лучшую известную реализацию на PBC на одно вычислительное ядро с рабочей частотой в 1,25 раза выше, с сохранением оптимальной степени сжатия методом Хаффмана для текущего сообщения.

Разработанная модификация позволяет организовать сжатие плотного потока данных на скорости до 128 Гбит/с, что обеспечивает потребности современных интерфейсов, таких как PCI Express 3.0, Thunderbolt 3, Ethernet 100GbE, InfiniBand EDR 4X. При этом масштабирование вычислительного блока позволит обрабатывать потоки данных на большей скорости.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК


1. *Berisha B., Mëziu E., Shabani I.* Big data analytics in Cloud computing: an overview // Journal of Cloud Computing. – 2022. – No. 11 (1).
2. *Morgan S.* The 2020 Data Attack Surface Report, White Paper. – https://cybersecurityventures.com/wp-content/uploads/2020/12/ArcserveDataReport2020.pdf.
3. 2020: Oracle’s Top 10 Cloud Predictions. – https://www.oracle.com/a/ocom/docs/cloud/oracle-cloud-predictions-2020.pdf 2020г.
4. *Попов А.В.* Алгоритмы энтропийного кодирования при сжатии спектра телевизионного сигнала // T-Comm — Телекоммуникации и Транспорт. – 2013. – № 4. – С. 43-46.
5. *Луканов А.С.* Системы реального времени: учеб. пособие. – Самара: Изд-во Самарского университета, 2020. – 156 с.
6. https://aws.amazon.com/ru/blogs/aws/data-compression-improvements-in-amazon-redshift/.
7. *Promberger L., Schwemmer R., Fröning H*. Characterization of data compression across CPU platforms and accelerators // Concurrency and Computation: Practice and Experience. – 2021. – Vol. 35, No. 20 (e6465).
8. *Knorr F., Thoman P., Fahringer T.* ndzip-gpu: Efficient Lossless Compression of Scientific Floating-Point Data on GPUs // SC '21: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis Article. – Nov. 2021. – No. 93. – P. 1-14.
9. *Deaton J.D., Purdy J., Bacon* A. Smashing Big Data Costs with GZIP Hardware, AHA Products Group, Technical White Paper.
10. *Qiao W., Du J., Fang Z., Lo M. et. al.* High-Throughput Lossless Compression on Tightly Coupled CPU-FPGA Platforms // 2018 IEEE 26th Annual International Symposium on Field-Programmable Custom Computing Machines. – 2018. – P. 37-44.
11. *Huffman D.A.* A method for the construction of minimum-redundancy codes // Proceedings of the IRE. – Sept. 1952. – Vol. 40, No. 9. – P. 1098-1101.
12. *Shannon C.E.* A mathematical theory of communication // Bell Sys. Tech. Jour. – 1948. – Vol. 27. – P. 398-403.
13. *Witten I.H., Moffat A., Bell T.C.* Managing Gigabytes: Compressing and Indexing Documents and Images. – 1st ed. – Van Nostrand Reinhold, New York, USA, 1994. – 429 p.
14. *Moffat A., Turpin A.* On the Implementation of Minimum Redundancy Prefix Codes // IEEE Transactions on Communications. – 1997. – Vol. 45, No. 10. – P. 1200-1207.
15. *Hashemian R.* Memory Efficient and High-speed Search Huffman Coding // IEEE Transactions on Communications. – 1995. – Vol. 43, No. 10. – P. 2576-2581.

16. *Kraft L.G*. A Device for Quantizing, Grouping, and Coding Amplitude-modulated Pulses. – Massachusetts Institute of Technology. Dept. of Electrical Engineering, 1949.
17. International Standard ISO/IEC 10918-1:1993(E). CCITT Rec. T.81 (1992 E). Information technology — Digital compression and coding of continuous-tone still images — Requirements and Guidelines.
18. *Каляев А.В., Левин И.И.* Модульно-наращиваемые многопроцессорные системы со структурно-процедурной организацией вычислений. – М.: Янус-К, 2003. – 380 с.
19. *Каляев И.А., Левин И.И.* Реконфигурируемые вычислительные системы на основе ПЛИС. – Ростов-на-Дону: Изд-во ЮФУ, 2021. – 458 с.
20. *Alekseev K.N., Sorokin D.A., Levin I.I.* Structural-procedural implementation the huffman coding on reconfigurable computer systems in real time // Vestnik Komp'iuternykh i Informatsionnykh Tekhnologii (Herald of Computer and Information Technologies). – 2018. – No. 9. – P. 3-10.

REFERENCES

1. *Berisha B., Mëziu E., Shabani I.* Big data analytics in Cloud computing: an overview, *Journal of Cloud Computing,* 2022, No. 11 (1).
2. *Morgan S.* The 2020 Data Attack Surface Report, White Paper. Available at: https://cybersecurityventures.com/wp-content/uploads/2020/12/ArcserveDataReport2020.pdf.
3. 2020: Oracle's Top 10 Cloud Predictions. Available at: https://www.oracle.com/a/ocom/docs/cloud/oracle-cloud-predictions-2020.pdf 2020.
4. *Popov A.V.* Algoritmy entropiynogo kodirovaniya pri szhatii spektra televizionnogo signala [Entropy coding algorithms for spectrum compression of television signals], *T-Comm — Telekommunikatsii i Transport* [T-Comm — Telecommunications and Transport], 2013, No. 4, pp. 43-46.
5. *Lukanov A.S.* Sistemy real'nogo vremeni: ucheb. posobie [Real-time systems: a tutorial]. Samara: Izd-vo Samarskogo universiteta, 2020, 156 p.
6. Available at: https://aws.amazon.com/ru/blogs/aws/data-compression-improvements-in-amazon-redshift/.
7. *Promberger L., Schwemmer R., Fröning H*. Characterization of data compression across CPU platforms and accelerators, *Concurrency and Computation: Practice and Experience,* 2021, Vol. 35, No. 20 (e6465).
8. *Knorr F., Thoman P., Fahringer T.* ndzip-gpu: Efficient Lossless Compression of Scientific Floating-Point Data on GPUs, *SC '21: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis Article,* Nov. 2021, No. 93, pp. 1-14.
9. *Deaton J.D., Purdy J., Bacon* A. Smashing Big Data Costs with GZIP Hardware, AHA Products Group, Technical White Paper.
10. *Qiao W., Du J., Fang Z., Lo M. et. al.* High-Throughput Lossless Compression on Tightly Coupled CPU-FPGA Platforms, *2018 IEEE 26th Annual International Symposium on Field-Programmable Custom Computing Machines,* 2018, pp. 37-44.
11. *Huffman D.A.* A method for the construction of minimum-redundancy codes, *Proceedings of the IRE,* Sept. 1952, Vol. 40, No. 9, pp. 1098-1101.
12. *Shannon C.E.* A mathematical theory of communication, *Bell Sys. Tech. Jour*., 1948, Vol. 27, pp. 398-403.
13. *Witten I.H., Moffat A., Bell T.C.* Managing Gigabytes: Compressing and Indexing Documents and Images. 1st ed. Van Nostrand Reinhold, New York, USA, 1994, 429 p.
14. *Moffat A., Turpin A.* On the Implementation of Minimum Redundancy Prefix Codes, *IEEE Transactions on Communications,* 1997, Vol. 45, No. 10, pp. 1200-1207.
15. *Hashemian R.* Memory Efficient and High-speed Search Huffman Coding, *IEEE Transactions on Communications,* 1995, Vol. 43, No. 10, pp. 2576-2581.
16. *Kraft L.G*. A Device for Quantizing, Grouping, and Coding Amplitude-modulated Pulses. Massachusetts Institute of Technology. Dept. of Electrical Engineering, 1949.
17. International Standard ISO/IEC 10918-1:1993(E). CCITT Rec. T.81 (1992 E). Information technology — Digital compression and coding of continuous-tone still images — Requirements and Guidelines.
18. *Kalyaev A.V., Levin I.I.* Modul'no-narashchivaemye mnogoprotsessornye sistemy so strukturno-protsedurnoy organizatsiey vychisleniy [Modular-scalable multiprocessor systems with structural-procedural organization of computations]. Moscow: Yanus-K, 2003, 380 p.
19. *Kalyaev I.A., Levin I.I.* Rekonfiguriruemye vychislitel'nye sistemy na osnove PLIS [Reconfigurable computing systems based on FPGAs]. Rostov-on-Don: Izd-vo YuFU, 2021, 458 p.
20. *Alekseev K.N., Sorokin D.A., Levin I.I.* Structural-procedural implementation the huffman coding on reconfigurable computer systems in real time, *Vestnik Komp'iuternykh i Informatsionnykh Tekhnologii (Herald of Computer and Information Technologies),* 2018, No. 9, pp. 3-10.

Статью рекомендовал к опубликованию д.т.н., профессор Ю.А. Кравченко.

**Левин Илья Израилевич** – Южный федеральный университет; e-mail: levin@superevm.ru; г. Таганрог, Россия; тел.: +78634612111; кафедра интеллектуальных и многопроцессорных систем; зав. кафедрой; д.т.н.; профессор.

**Дудников Евгений Александрович** – e-mail: everlast-83@mail.ru; тел.: +79185914907; кафедра интеллектуальных и многопроцессорных систем; аспирант.

**Levin Ilya Izrailevich** – Southern Federal University; e-mail: levin@superevm.ru; Taganrog, Russia; phone: +78634612111; the Department of Intellectual and Multiprocessor Systems; head of the department; dr. of eng. sc.; professor.

**Dudnikov Evgeny Alexandrovich** – e-mail: everlast-83@mail.ru; phone: +79185914907; the Department of Intellectual and Multiprocessor Systems; graduate student.

УДК 004.056 DOI 10.18522/2311-3103-2024-5-58-68

**С.Ю. Мельников, Р.В. Мещеряков, В.А. Пересыпкин**

## НЕКОТОРЫЕ АСПЕКТЫ ПРИМЕНЕНИЯ ТЕХНОЛОГИЙ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В ЗАДАЧАХ ЗАЩИТЫ ИНФОРМАЦИИ (ОБЗОР)

*Технологии искусственного интеллекта (ИИ) являются одной из наиболее динамично развивающихся областей обработки информации. Технологии ИИ используются как для обеспечения защиты информации, так и для организации атак на средства ее защиты. Сами системы ИИ могут содержать уязвимости и быть подвержены атакам различного рода. В статье анализируются некоторые аспекты применения технологий ИИ в задачах защиты информации. В рамках задачи биометрической идентификации рассматриваются угрозы подделки биометрических идентификационных признаков с целью получения прав доступа, и способы противодействия таким угрозам. Анализируются преимущества использования ИИ при защите информации в компьютерных системах и сетях по сравнению с традиционными средствами защиты. На примере акустического канала утечки информации от клавиатуры иллюстрируется использование технологий ИИ для обработки данных из технических каналов утечки. Рассматриваются методы повышения информативности таких каналов, использующие временные сверточные сети и модели классификации изображений, а также способы противодействия им. Отдельное внимание уделено вопросам информационной безопасности в набирающих популярность системах сжатия и передачи информации без значительных смысловых потерь (направление Semantic Communications). Рассматриваются ряд вопросов информационной безопасности, возникающих при использовании больших языковых моделей типа ChatGPT, способных массово генерировать уникальный «человекоподобный» контент и использовать его для организации фишинговых и других атак социальной инженерии. Описана атака на системы ИИ с использованием скрытого канала. Уделено внимание необходимости развития технологий доверенного искусственного интеллекта.*

*Информационная безопасность; кибербезопасность; технический канал утечки; искусственный интеллект; доверенный искусственный интеллект.*

**S.Yu. Melnikov, R.V. Meshcheryakov, V.A. Peresypkin**

## SOME ASPECTS OF APPLICATION OF ARTIFICIAL INTELLIGENCE TECHNOLOGIES IN INFORMATION SECURITY (REVIEW)

*Artificial intelligence (AI) technologies are one of the most dynamically developing areas of information processing. AI technologies are used both to ensure the information security and to organize attacks on information security tools. AI systems themselves may contain vulnerabilities and be susceptible to various types of attacks. The article analyzes some aspects of the use of AI technologies in information security tasks. Within the framework of the task of biometric identification, threats of falsification of biometric identification characteristics in order to obtain access rights, and ways to counter such threats are considered. The advantages of using AI in protecting information in computer systems and networks in comparison with traditional means of protection are analyzed. Using the example of an acoustic channel of information leakage from a keyboard, the use of AI technologies for processing data from technical leakage channels is illustrated. Methods for increasing the information content of such channels using*

*temporary convolutional networks and image classification models, as well as ways to counter them, are considered. Special attention is paid to information security issues in increasingly popular systems for compressing and transmitting information without significant semantic losses (Semantic Communications). A number of information security issues that arise when using large language models such as ChatGPT, capable of massively generating unique "human-like" content and using it to organize phishing and other social engineering attacks, are considered. An attack on AI systems using a covert channel is described. Attention is paid to the need to develop trusted artificial intelligence technologies.*

*Information security; cybersecurity; technical leakage channel; artificial intelligence; trusted artificial intelligence.*

**Введение.** Под искусственным интеллектом понимается комплекс технологических решений, позволяющий имитировать когнитивные функции человека и получать при выполнении конкретных задач результаты, сопоставимые с результатами интеллектуальной деятельности человека или превосходящие их [1].

Технологии ИИ являются динамично развивающийся областью обработки информации, включающей в себя обработку естественного языка, компьютерное зрение, распознавание и синтез речи, интеллектуальную поддержку принятия решений и др. К основным задачам, которые могут решаться с использованием методов ИИ, относятся: классификация, кластеризация, распознавание образов, обнаружение аномалий, прогнозирование, обработка естественного языка, инженерия знаний, создание экспертных систем.

Стремительное развитие технологий ИИ, которые используются как для обеспечения защиты информации, так и для организации атак на средства ее защиты, в ближайшее время приведет к значительным изменениям ландшафта индустрии информационной безопасности. Кроме того, сами системы ИИ могут содержать уязвимости и быть подвержены атакам различного рода.

**Угрозы безопасности информации, создаваемые с использованием технологий ИИ, и способы противодействия им**

1. Первые методы ИИ разрабатывались для решения идентификационных задач, связанных с обработкой речи и текста [2]. В настоящее время биометрическая идентификация используется для управления доступом, например, в современных мобильных телефонах и планшетах.

К классическим угрозам, реализуемым с использованием ИИ, можно отнести [3] подделку биометрических идентификационных признаков с целью получения прав доступа путем формирования идентификационных признаков другого лица. К возможным современным угрозам следует отнести формирование ложных видео/речевых/текстовых сообщений, имитирующих конкретного человека; создание ложных фото и видео с участием конкретных лиц; подделку почерка; подделку авторского стиля текстов и др. Способы создания фальшивого контента приведены на рис. 1, результаты их применения – на рис. 2 [4].


*Рис. 1. Способы создания фальшивого контента.*

Угрозы биометрическим системам с использованием подходов на основе ИИ предполагают нарушение не только конфиденциальности и доступности информации, но и ее целостности, т.к. может производиться подмена информации. Помимо подделки внешних идентификационных признаков, которые различимы, возможна подделка более сложных «внутренних» признаков [4], связанных с социальными и иными поведенческими реакциями (рис. 2).

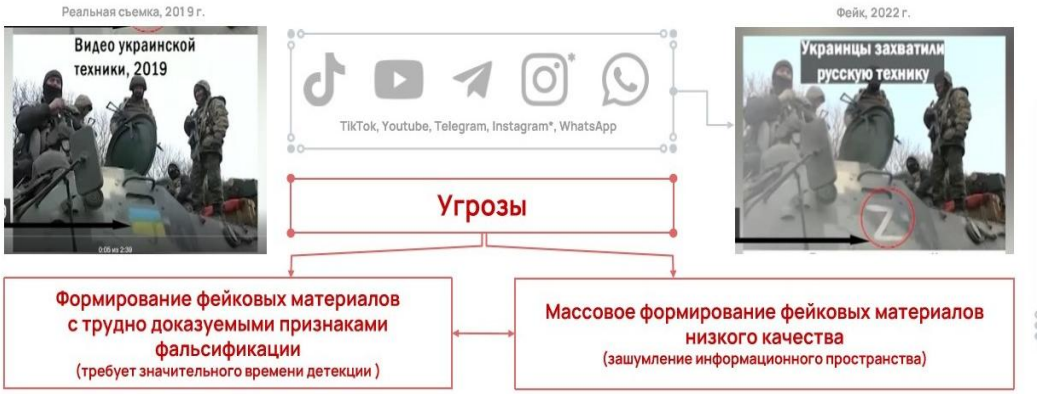



*Рис. 2. Иллюстрация результатов применения технологии deepfake*

В связи с вышесказанным возникают задачи противодействия угрозам безопасности информации с использованием технологий ИИ [5]. К ним, в частности, относятся: выявление источника угроз, который использует технологии ИИ (рис. 3) [4]; классификация угроз в отношении защищаемого объекта и формирование нового пространства признаков угроз; моделирование действий злоумышленника, в том числе с помощью поведенческого анализа [6]; формирование стратегии защиты для выявления аномалий, генерируемых системами с ИИ; разработка новых методов защиты от атак с использованием генеративных моделей ИИ, в том числе больших мультимодальных языковых моделей.


*Рис. 3. Основные способы выявления deepfake*

2. К основным задачам защиты информации в компьютерных системах и сетях с использованием ИИ относятся [5]: обнаружение компьютерных атак и вредоносных программ; обнаружение модификаций данных; предотвращение утечек конфиденциальных данных по тем или иным техническим каналам; повышение надежности и киберустойчивости компьютерных систем и сетей.

Недостатки традиционных систем информационной безопасности во многом связаны с тем, что они используют заблаговременно сформированные подходы к выявлению угроз и варианты реакций на них. Это влечет за собой как неспособность быстро реагировать на новые угрозы, так и появление большого числа ложных срабатываний. Другим важным ограничением являются объемы анализируемых данных. Существующие системы могут генерировать значительное количество показателей, связанных с событиями информационной безопасности, их вычислительно сложно анализировать в реальном времени.

К важнейшим преимуществам технологий ИИ следует отнести [7, 8]:

- ♦ возможность быстрой обработки больших массивов данных для раннего предупреждения о критических событиях безопасности;
- ♦ возможность одновременного анализа данных из нескольких источников, включая сетевой трафик, системные журналы и данные о поведении пользователей, чтобы выявлять подозрительные аномалии;
- ♦ возможность автоматического реагирования на угрозы, включая автоматический запрет доступа к скомпрометированной системе с целью предотвращения дальнейшего ущерба.

Средства интеллектуализации также позволяют провести моделирование различных угроз и поведения нарушителя периметра безопасности систем контроля управления доступом (СКУД), а также обеспечить на рубежах охраны средства контроля с использованием распознавания видеобразов пользователей СКУД.

3. С позиций нападающей стороны перспективным представляется использование технологий ИИ для обработки данных из технических каналов утечки информации [9]. Технологии ИИ могут использоваться как для «проигрывания» различных сценариев кибератак с использованием технических каналов утечки, так и для обработки (распознавания) сигналов, регистрируемых в этих каналах. На рис. 4. представлен фрагмент сигнала, полученного по акустическому каналу от клавиатуры компьютера [10]. Использование технологий ИИ позволяет значимо повысить качество классификации фрагментов сигнала, соответствующих нажатиям на клавиши, обеспечив практически однозначное определение истинных клавиш.

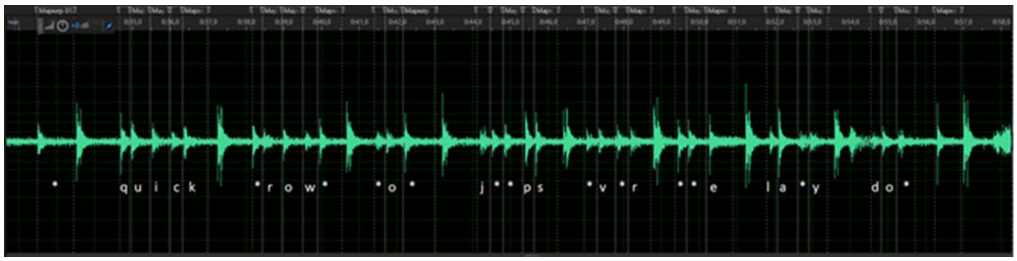



*Рис. 4. Результаты нейросетевого распознавания фрагментов сигнала*

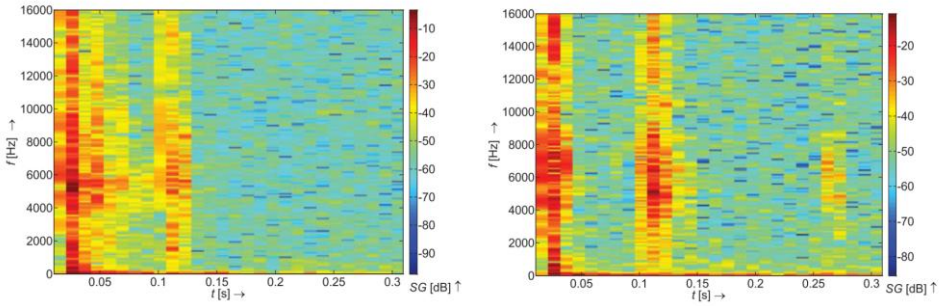



*Рис. 5. Спектрограммы звуков нажатия клавиш «L» и «D» ([11])*

На графиках на рис. 5 представлены спектрограммы звуков нажатий на клавиши «L» и «D». Ось абсцисс соответствует времени (в секундах), ось ординат – частоте (в герцах), а цвета соответствуют амплитудам наблюдаемых частот в данный момент времени (яркие цвета обозначают большие амплитуды). Эту спектрограмму можно рассматривать как изображение, что сводит [11] проблему распознавания звуков отдельных клавиш к проблеме классификации изображений. В [12] использовалась модель глубокого обучения для распознавания нажатий клавиш на ноутбуке. Звуки нажатия клавиш фиксировались с помощью встроенного микрофона рядом расположенного смартфона. В качестве входных данных для модели классификации изображений CoAtNet использовались мел-спектрограммы.

Еще один подход с использованием временных сверточных сетей (Temporal Convolutional Network, TCN) изложен в [13]. Предложенный классификатор достиг точности 95% без использования языковой модели. Для постобработки, то есть коррекции ошибок, применяются те или иные вероятностные модели текста [10, 14], которые могут значительно повысить точность распознавания.

В работе [15] предложен способ добавления специальных «фальшивых» звуков нажатия клавиш для противодействия рассматриваемым атакам.

4. Развитие ИИ привело к появлению принципиально новых технологий сжатия и передачи информации без значительных смысловых потерь (направление Semantic Communications) [16]. В свою очередь, расширяющееся использование таких технологий в системах связи и прежде всего, в сетях интернета вещей (IoT) [17], делает необходимым проведение тщательного анализа возникающих задач в области информационной безопасности. Семантическая коммуникация направлена на адресную и точную передачу смысла между различными интеллектуальными агентами, как людьми, так и машинами. В отличие от обычных систем связи, в которых приоритетна точность передачи исходящих данных, при семантической коммуникации приоритет принадлежит семантической точности. Семантическая коммуникация делает упор на извлечение и передачу информации семантического уровня из сообщения с целью «сохранения смысла». Современные коммуникационные технологии направлены на передачу большего количества данных с меньшим количеством ошибок с минимумом затрат (в рамках шенноновской парадигмы), тогда как семантическая коммуникация стремится передать максимальный объем семантики с наименьшими затратами коммуникационных ресурсов, быстро доставляя семантический контент (уже вне шенонновской парадигмы). В [18], например, для извлечения смысла передаваемого текста использовалась архитектура Transformer. Для извлечения смысла при передаче изображений в [19] использована сверточная нейронная сеть. Однако по мере развития таких систем они сталкиваются со значительными проблемами безопасности, конфиденциальности и доверия при интеграции технологий ИИ в интеллектуальные коммуникационные приложения. Перехват или зашумление сообщений при семантической коммуникации создает более серьезные угрозы безопасности [20], чем в обычных системах передачи информации на уровне бит или символов. Разрабатываются новые подходы к безопасности таких систем, в том числе использующие физический уровень канала (Physical Layer Security, PLS) [21], предлагаются метрики для оценки безопасности семантических коммуникаций [22].

5. Технологии ИИ могут быть использованы не только в технических областях. Использование больших языковых моделей типа GPT позволяет повысить эффективность деятельности аналитических подразделений организаций и предприятий. Спектр применения таких моделей весьма широк и позволяет не только совершенствовать аналитическую деятельность и конкурентную разведку, но также массово генерировать уникальный «человекоподобный» контент для проведения атак социальной инженерии [23, 24]. В последнее время широкое распространение больших языковых моделей приводит к росту социально-культурных рисков для гражданского общества [25].

Растущие возможности систем генеративного ИИ, таких как ChatGPT, привели к увеличению генерации синтетического контента, что имеет последствия для различных секторов общественной жизни, включая средства массовой информации, кибербезопас-

ность, образование, социальные сети, художественное творчество и др. В [26] проведен мониторинг основных мировых новостных веб-сайтов, и показано, что за последние полтора года доля новостных публикаций, созданных генеративным ИИ, удвоилась. Особую тревогу у авторов вызывает рост числа дезинформирующих и фейковых публикаций. Отметим однако, что существует и альтернативная точка зрения ([27]), согласно которой распространенные представления об опасности генеративного ИИ для производства фейков и дезинформации являются преувеличенными.

Актуальной становится проблема обнаружения контента, созданного LLM. Одной из новых и важных задач в области обработки текстов сейчас является определение того, написан ли данный текст человеком или генеративным ИИ. Это задача бинарной классификации, в которой анализируемый текст должен быть отнесен к одному из двух классов: тексты, имеющие естественное или искусственное происхождение. Для человека эта задача оказывается весьма сложной. Так, в экспериментах, проведенных в [28], точность решений, принимаемых экспертами-людьми, составила всего 61%. Однако эта задача хорошо решается алгоритмически, точность современных методов детектирования машинно-сгенерированных текстов достигает 90% и более [29–31].

**Обеспечение безопасности систем искусственного интеллекта.** Отдельным направлением обеспечения информационной безопасности является обеспечение безопасности самих систем ИИ. Новыми вызовами для государства, как отмечается в Национальной стратегии развития ИИ, являются, в том числе, «возникновение в сфере разработки, создания и использования ИИ новых типов угроз информационной безопасности, нехарактерных для других сфер применения информационных технологий». Технологии ИИ обладают той особенностью, что алгоритм решения задачи не фиксирован заранее, а формируется в процессе ее решения и существенным образом зависит от входных данных (обучающей выборки). Это приводит [32] к новым возможным атакам на системы ИИ, таким, как атаки на обучающие данные, искажение разметки, атаки, направленные на установление принадлежности конкретных данных обучающей выборке, атаки, направленные на получение данных из обученной модели, атаки на уровне вычислительных платформ и др. (рис. 6) [4].

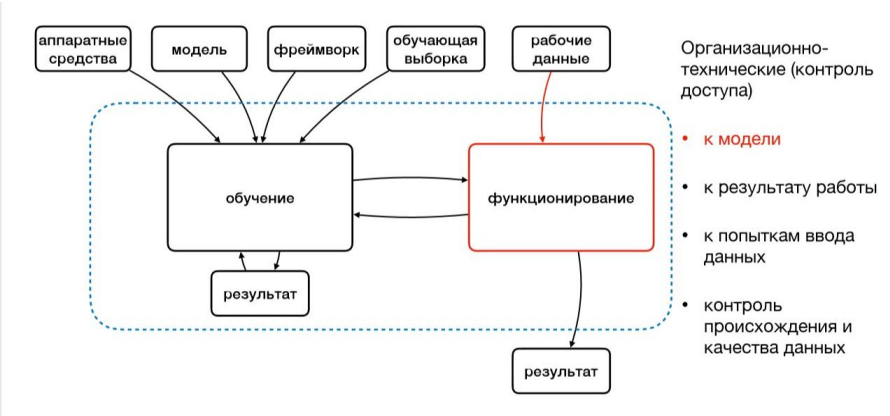



*Рис. 6. Организационно-технические средства защиты систем ИИ*

При построении информационных систем возможны атаки с помощью т.н. «скрытых каналов» [33]. Отметим недавно предложенную атаку [34] с использованием скрытого канала, напоминающего акустический технический канал утечки от клавиатуры. Атака направлена на приложения (помощники) с искусственным интеллектом, которые все интенсивнее проникают в нашу жизнь, их используют для получения совета или помощи в личных и конфиденциальных вопросах. В качестве скрытого канала рассматривается просто длина токенов сообщения. Информационный обмен пользователя с LLM, которая выполняется на удаленном сервере, защищен, и осуществляется в зашифрованном виде.

Однако LLM генерирует и отправляет ответ в виде серии токенов, причем каждый токен передается по мере его создания. Несмотря на шифрование, размер пакетов может раскрыть длины использованных токенов. Определение содержания ответа только на основе последовательности длин токенов, конечно, является непростой задачей и может допускать несколько вариантов решения. Однако, используя специально настроенную для решения такой задачи LLM, злоумышленник в ряде случаев может восстановить передаваемые тексты. Доля успешных атак на ChatGPT-4 от OpenAI и Copilot от Microsoft с помощью такого подхода составила от 17 до 53%.

Ряд систем генеративного искусственного интеллекта (рис. 7) [4] подвержена не только “переобучениям”, но и “галлюцинациям” – выдается по запросу информация, которой нет в обучающей выборке. Указанная уязвимость может быть эксплуатируема злоумышленниками, что еще раз подчеркивает актуальность задачи создания доверенного искусственного интеллекта [35].

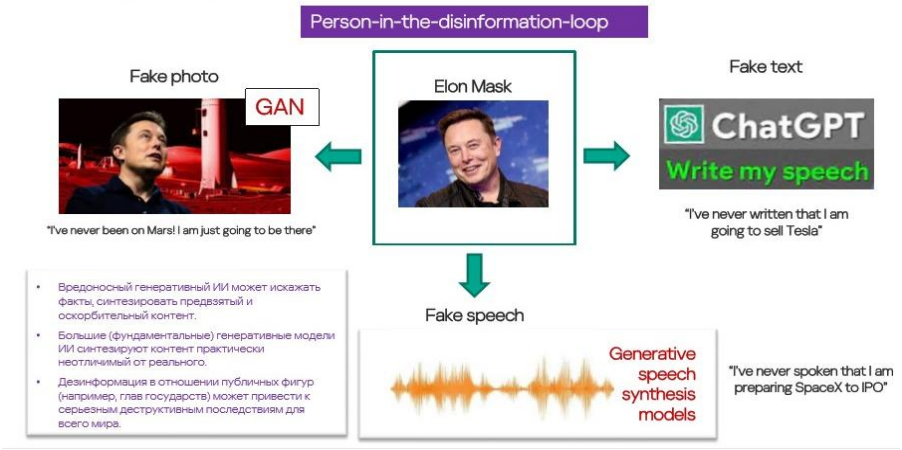



*Рис. 7. Пример системы генеративного ИИ для дискредитации личности*

Для защиты систем ИИ наряду с типовыми средствами защиты информации должны использоваться и специализированные технологии, и средства защиты, к основным из которых можно отнести: повышение надежности обучающих выборок, оценка доверия к принимаемым решениям, интерпретируемость результатов, контроль процессов обучения и верификации, повторяемость, отсутствие галлюцинаций и др. Основным направлением здесь должно стать создание систем доверенного искусственного интеллекта.

**Заключение.** Обеспечение высокого уровня информационной безопасности сегодня требует привлечения широкого спектра технологий ИИ. Полноценно владеющий этими технологиями будет превосходить противника вне зависимости от выполняемой функции - атакующей или нападающий. Развитие технологий ИИ для обработки данных различной природы делает необходимым формирование новых требований к моделям угроз и к средствам защиты информации. Основным направлением защиты собственно систем ИИ должно стать развитие технологий доверенного искусственного интеллекта.

*Работа выполнена при финансовой поддержке гранта РНФ № 24-11-00340.*

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Национальная стратегия развития искусственного интеллекта на период до 2030 года, утверждена Указом Президента Российской Федерации от 10 октября 2019 г. № 490.
2. *McCorduck P., Cfe C.* Machines who think: A personal inquiry into the history and prospects of artificial intelligence. – AK Peters/CRC Press, 2004.
3. *Mughal A.A.* Artificial Intelligence in Information Security: Exploring the Advantages, Challenges, and Future Directions // Journal of Artificial Intelligence and Machine Learning in Management. – 2018. – Vol. 2, No. 1. – P. 22-34.

4. Материалы первого форума «Цифровая экономика. Технологии доверенного искусственного интеллекта». Москва, МГУ, кластер «Ломоносов», 23 мая 2023 г. – Режим доступа: https://ib-bank.ru/trust-ai/materials23 (дата обращения: 10.10.2024).
5. *Мещеряков Р.В., Мельников С.Ю., Пересыпкин В.А., Хорев А.А*. Перспективные направления применения технологий искусственного интеллекта при защите информации // Вопросы кибербезопасности. – 2024. – № 4 (62). – С. 2-12. – DOI: 10.21681/2311-3456-2024-4-02-12. – EDN GJWQWP.
6. *Shelke P., Hämäläinen T.* Analysing Multidimensional Strategies for Cyber Threat Detection in Security Monitoring / In M. Lehto, & M. Karjalainen (Eds.) // Proceedings of the 23rd European Conference on Cyber Warfare and Security. – 2024. – 23. – P. 780-787. Academic Conferences International Ltd. Proceedings of the European Conference on Cyber Warfare and Security. – Режим доступа: https://doi.org/10.34190/eccws.23.1.2123 (дата обращения: 10.10.2024).
7. *Аветисян А.И.* Кибербезопасность в контексте искусственного интеллекта // Вестник Российской академии наук. – 2022. – Т. 92, № 12. – С. 1119-1123. – DOI: 10.31857/S0869587322120039. – EDN RYZRRU.
8. *Camacho N.G.* The Role of AI in Cybersecurity: Addressing Threats in the Digital Age // Journal of Artificial Intelligence General science (JAIGS). – 2024. – Vol. 3, No. 1. – P. 143-154. – ISSN: 3006-4023.
9. *Panoff M. et al.* A review and comparison of AI-enhanced side channel analysis // ACM Journal on Emerging Technologies in Computing Systems (JETC). – 2022. – Vol. 18, No. 3. – P. 1-20.
10. *Zhuang L., Zhou F., Tygar J.D.* Keyboard acoustic emanations revisited // ACM Transactions on Information and System Security (TISSEC). – 2009. – Vol. 13, No. 1. – P. 1-26.
11. *Taheritajar A., Harris Z. M., Rahaeimehr R.* A Survey on Acoustic Side Channel Attacks on Keyboards // arXiv preprint arXiv:2309.11012. – 2023.
12. *Harrison J., Toreini E., Mehrnezhad M.* A practical deep learning-based acoustic side channel attack on keyboards //2023 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW). – IEEE, 2023. – P. 270-280.
13. *Spata M. O. et al.* A New Deep Learning Pipeline for Acoustic Attack on Keyboards // IntelliSys 2024. – Cham: Springer Nature Switzerland, 2024. – P. 402-414.
14. *Мельников С.Ю., Пересыпкин В.А*. Об эволюции классических вероятностных моделей языка в естественно-языковых приложениях // Вестник современных цифровых технологий. – 2023. – № 16. – С. 4-14. – EDN YDIGDT.
15. *Rodrigues D. et al.* A Prototype for Generating Random Key Sounds to Prevent Keyboard Acoustic Side-Channel Attacks // 2024 IEEE 22nd Mediterranean Electrotechnical Conference (MELECON). – IEEE, 2024. – P. 1287-1292.
16. *Yang W. et al.* Semantic communications for future internet: Fundamentals, applications, and challenges // IEEE Communications Surveys & Tutorials. – 2022. – Vol. 25, No. 1. – P. 213-250.
17. *Wang Y.* Semantic Communication Networks Empowered Artificial Intelligence of Things // 2024 IEEE Annual Congress on Artificial Intelligence of Things (AIoT). – IEEE, 2024. – P. 189-193.
18. *Xie H. et al.* Deep learning enabled semantic communication systems // IEEE Transactions on Signal Processing. – 2021. – Vol. 69. – P. 2663-2675.
19. *Bourtsoulatze E., Kurka D. B., Gündüz D.* Deep joint source-channel coding for wireless image transmission // IEEE Transactions on Cognitive Communications and Networking. – 2019. – Vol. 5, No. 3. – P. 567-579.
20. *Luo X. et al.* Encrypted semantic communication using adversarial training for privacy preserving // IEEE Communications Letters. – 2023. – Vol. 27, No. 6. – P. 1486-1490.
21. *Nguyen V.L. et al.* Security and privacy for 6G: A survey on prospective technologies and challenges // IEEE Communications Surveys & Tutorials. – 2021. – Vol. 23, No. 4. – P. 2384-2428.
22. *Li Y. et al.* Secure Semantic Communications: From Perspective of Physical Layer Security // IEEE Communications Letters. – 2024. – DOI: 10.1109/LCOMM.2024.3452715.
23. *Hazell J.* Large language models can be used to effectively scale spear phishing campaigns // arXiv preprint arXiv:2305.06972. – 2023.
24. *Greco F. et al.* David versus Goliath: Can Machine Learning Detect LLM-Generated Text? A Case Study in the Detection of Phishing Emails // ITASEC 2024: The Italian Conference on CyberSecurity, Italy. CEUR-WS Vol. 3731. – 2024.
25. *Былевский П.Г.* Социально-культурные риски мультимодальных больших генеративных моделей «искусственного интеллекта» (GenAI) // Культура и искусство. – 2024. – № 6. – С. 213-224. – DOI: 10.7256/2454-0625.2024.6.70926. – EDN: DWMERQ.
26. *Hanley H.W.A., Durumeric Z.* Machine-made media: Monitoring the mobilization of machine-generated articles on misinformation and mainstream news websites // Proceedings of the International AAAI Conference on Web and Social Media. – 2024. – Vol. 18. – P. 542-556.

27. *Simon F.M., Altay S., Mercier H.* Misinformation reloaded? Fears about the impact of generative AI on misinformation are overblown // Harvard Kennedy School Misinformation Review. – 2023. – Vol. 4, No. 5.
28. *Liu Y. et al.* ArguGPT: evaluating, understanding and identifying argumentative essays generated by GPT models // arXiv preprint arXiv:2304.07666. – 2023.
29. *Wu J. et al.* A survey on llm-gernerated text detection: Necessity, methods, and future directions // arXiv preprint arXiv:2310.14724. – 2023.
30. *Ghosal S.S. et al.* A Survey on the Possibilities & Impossibilities of AI-generated Text Detection // Transactions on Machine Learning Research. – No. 1. – 2024.
31. *Sadasivan V.S. et al.* Can AI-generated text be reliably detected? // arXiv preprint arXiv:2303.11156. – 2023.
32. *Маршалко Г.Б., Романенков Р.А., Труфанова Ю.А*. Анализ безопасности проекта национального стандарта «Нейросетевые алгоритмы в защищенном исполнении. Автоматическое обучение нейросетевых моделей на малых выборках в задачах классификации» // Тр. Института системного программирования РАН. – 2023. – Т. 35, № 6. – С. 179-188. – DOI: 10.15514/ISPRAS-2023-35(6)-11. – EDN HNDIYD.
33. *Грушо А.А.* Скрытые каналы и безопасность информации в компьютерных системах // Дискретная математика. – 1998. – Т. 10, № 1. – С. 3-9.
34. *Weiss R. et al.* What Was Your Prompt? A Remote Keylogging Attack on AI Assistants // arXiv preprint arXiv:2403.09751. – 2024.
35. *Турдаков Д.Ю., Аветисян А.И., Архипенко К.В. [и др.]*. Доверенный Искусственный интеллект: вызовы и перспективные решения // Доклады Российской академии наук. Математика, информатика, процессы управления. – 2022. – Т. 508, № 1. – С. 13-18. – DOI: 10.31857/S2686954322070207. – EDN CVIVCS.

REFERENCES

1. Natsional'naya strategiya razvitiya iskusstvennogo intellekta na period do 2030 goda, utverzhdena Ukazom Prezidenta Rossiyskoy Federatsii ot 10 oktyabrya 2019 g. № 490 [National Strategy for the Development of Artificial Intelligence through 2030, approved by the Decree of the President of the Russian Federation, October 10, 2019, No. 490].
2. *McCorduck P., Cfe C.* Machines who think: A personal inquiry into the history and prospects of artificial intelligence. AK Peters/CRC Press, 2004.
3. *Mughal A.A.* Artificial Intelligence in Information Security: Exploring the Advantages, Challenges, and Future Directions, *Journal of Artificial Intelligence and Machine Learning in Management,* 2018, Vol. 2, No. 1, pp. 22-34.
4. Materialy pervogo foruma «Tsifrovaya ekonomika. Tekhnologii doverennogo iskusstvennogo intellekta». Moskva, MGU, klaster «Lomonosov», 23 maya 2023 g. [Proceedings of the first forum "Digital Economy. Technologies of Trusted Artificial Intelligence". Moscow, Moscow State University, Lomonosov Cluster, May 23, 2023]. Available at: https://ib-bank.ru/trust-ai/materials23 (accessed on 10 October 2024).
5. *Meshcheryakov R.V., Mel'nikov S.Yu., Peresypkin V.A., Khorev A.A.* Perspektivnye napravleniya primeneniya tekhnologiy iskusstvennogo intellekta pri zashchite informatsii [Promising areas of application of artificial intelligence technologies in information security], *Voprosy kiberbezopasnosti* [Cybersecurity Issues], 2024, No. 4 (62), pp. 2-12. DOI: 10.21681/2311-3456-2024-4-02-12. EDN GJWQWP.
6. *Shelke P., Hämäläinen T.* Analysing Multidimensional Strategies for Cyber Threat Detection in Security Monitoring, In M. Lehto, & M. Karjalainen (Eds.), *Proceedings of the 23rd European Conference on Cyber Warfare and Security,* 2024, 23, pp. 780-787. Academic Conferences International Ltd. Proceedings of the European Conference on Cyber Warfare and Security. Available at: https://doi.org/10.34190/eccws.23.1.2123 (accessed 10 October 2024).
7. *Avetisyan A.I.* Kiberbezopasnost' v kontekste iskusstvennogo intellekta [Cybersecurity in the Context of Artificial Intelligence],*Vestnik Rossiyskoy akademii nauk* [Bulletin of the Russian Academy of Sciences], 2022, Vo.. 92, No. 12, pp. 1119-1123. DOI: 10.31857/S0869587322120039. EDN RYZRRU.
8. *Camacho N.G.* The Role of AI in Cybersecurity: Addressing Threats in the Digital Age, *Journal of Artificial Intelligence General science (JAIGS*), 2024, Vol. 3, No. 1, pp. 143-154. ISSN 3006-4023.
9. *Panoff M. et al.* A review and comparison of AI-enhanced side channel analysis, *ACM Journal on Emerging Technologies in Computing Systems (JETC),* 2022, Vol. 18, No. 3, pp. 1-20.
10. *Zhuang L., Zhou F., Tygar J.D.* Keyboard acoustic emanations revisited, *ACM Transactions on Information and System Security (TISSEC),* 2009, Vol. 13, No. 1, pp. 1-26.

11. *Taheritajar A., Harris Z. M., Rahaeimehr R.* A Survey on Acoustic Side Channel Attacks on Keyboards, *arXiv preprint arXiv:2309.11012,* 2023.
12. *Harrison J., Toreini E., Mehrnezhad M.* A practical deep learning-based acoustic side channel attack on keyboards, *2023 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*. IEEE, 2023, pp. 270-280.
13. *Spata M. O. et al.* A New Deep Learning Pipeline for Acoustic Attack on Keyboards, *IntelliSys 2024.* Cham: Springer Nature Switzerland, 2024, pp. 402-414.
14. *Mel'nikov S.Yu., Peresypkin V.A.* Ob evolyutsii klassicheskikh veroyatnostnykh modeley yazyka v estestvenno-yazykovykh prilozheniyakh [On the evolution of classical probabilistic language models in natural language applications], *Vestnik sovremennykh tsifrovykh tekhnologiy* [Bulletin of modern digital technologies], 2023, No. 16, pp. 4-14. EDN YDIGDT.
15. *Rodrigues D. et al.* A Prototype for Generating Random Key Sounds to Prevent Keyboard Acoustic Side-Channel Attacks, *2024 IEEE 22nd Mediterranean Electrotechnical Conference (MELECON)*. IEEE, 2024, pp. 1287-1292.
16. *Yang W. et al.* Semantic communications for future internet: Fundamentals, applications, and challenges, *IEEE Communications Surveys & Tutorials,* 2022, Vol. 25, No. 1, pp. 213-250.
17. *Wang Y.* Semantic Communication Networks Empowered Artificial Intelligence of Things, *2024 IEEE Annual Congress on Artificial Intelligence of Things (AIoT)*. IEEE, 2024, pp. 189-193.
18. *Xie H. et al.* Deep learning enabled semantic communication systems, *IEEE Transactions on Signal Processing,* 2021, Vol. 69, pp. 2663-2675.
19. *Bourtsoulatze E., Kurka D. B., Gündüz D.* Deep joint source-channel coding for wireless image transmission, *IEEE Transactions on Cognitive Communications and Networking,* 2019, Vol. 5, No. 3, pp. 567-579.
20. *Luo X. et al.* Encrypted semantic communication using adversarial training for privacy preserving, *IEEE Communications Letters,* 2023, Vol. 27, No. 6, pp. 1486-1490.
21. *Nguyen V.L. et al.* Security and privacy for 6G: A survey on prospective technologies and challenges, *IEEE Communications Surveys & Tutorials,* 2021, Vol. 23, No. 4, pp. 2384-2428.
22. *Li Y. et al.* Secure Semantic Communications: From Perspective of Physical Layer Security, *IEEE Communications Letters,* 2024. DOI: 10.1109/LCOMM.2024.3452715.
23. *Hazell J.* Large language models can be used to effectively scale spear phishing campaigns, *arXiv preprint arXiv:2305.06972,* 2023.
24. *Greco F. et al.* David versus Goliath: Can Machine Learning Detect LLM-Generated Text? A Case Study in the Detection of Phishing Emails, *ITASEC 2024: The Italian Conference on CyberSecurity, Italy*. CEUR-WS Vol. 3731, 2024.
25. *Bylevskiy P.G.* Sotsial'no-kul'turnye riski mul'timodal'nykh bol'shikh generativnykh modeley «iskusstvennogo intellekta» (GenAI) [Socio-cultural risks of multimodal large generative models of "artificial intelligence" (GenAI)], *Kul'tura i iskusstvo* [Culture and Art], 2024, No. 6, pp. 213-224. DOI: 10.7256/2454-0625.2024.6.70926. EDN: DWMERQ.
26. *Hanley H.W.A., Durumeric Z.* Machine-made media: Monitoring the mobilization of machine-generated articles on misinformation and mainstream news websites, *Proceedings of the International AAAI Conference on Web and Social Media,* 2024, Vol. 18, pp. 542-556.
27. *Simon F.M., Altay S., Mercier H.* Misinformation reloaded? Fears about the impact of generative AI on misinformation are overblown, *Harvard Kennedy School Misinformation Review,* 2023, Vol. 4, No. 5.
28. *Liu Y. et al.* ArguGPT: evaluating, understanding and identifying argumentative essays generated by GPT models, *arXiv preprint arXiv:2304.07666,* 2023.
29. *Wu J. et al.* A survey on llm-gernerated text detection: Necessity, methods, and future directions, *arXiv preprint arXiv:2310.14724,* 2023.
30. *Ghosal S.S. et al.* A Survey on the Possibilities & Impossibilities of AI-generated Text Detection, *Transactions on Machine Learning Research,* No. 1*,* 2024.
31. *Sadasivan V.S. et al.* Can AI-generated text be reliably detected?, *arXiv preprint arXiv:2303.11156,* 2023.
32. *Marshalko G.B., Romanenkov R.A., Trufanova Yu.A.* Analiz bezopasnosti proekta natsional'nogo standarta «Neyrosetevye algoritmy v zashchishchennom ispolnenii. Avtomaticheskoe obuchenie neyrosetevykh modeley na malykh vyborkakh v zadachakh klassifikatsii» [Security analysis of the draft national standard "Neural network algorithms in secure implementation. Automatic training of neural network models on small samples in classification problems"], *Tr. Instituta sistemnogo programmirovaniya RAN* [Proceedings of the Institute for System Programming of the Russian Academy of Sciences], 2023, Vol. 35, No. 6, pp. 179-188. DOI: 10.15514/ISPRAS-2023-35(6)-11. EDN HNDIYD.
33. *Grusho A.A.* Skrytye kanaly i bezopasnost' informatsii v komp'yuternykh sistemakh [Covert channels and information security in computer systems], *Diskretnaya matematika* [Discrete Mathematics], 1998, Vol. 10, No. 1, pp. 3-9.

34. *Weiss R. et al.* What Was Your Prompt? A Remote Keylogging Attack on AI Assistants, *arXiv preprint arXiv:2403.09751,* 2024.
35. *Turdakov D.Yu., Avetisyan A.I., Arkhipenko K.V. [i dr.].* Doverennyy Iskusstvennyy intellekt: vyzovy i perspektivnye resheniya [Trusted Artificial Intelligence: Challenges and Promising Solutions], *Doklady Rossiyskoy akademii nauk. Matematika, informatika, protsessy upravleniya* [Reports of the Russian Academy of Sciences. Mathematics, informatics, control processes], 2022, Vol. 508, No. 1, pp. 13-18. DOI: 10.31857/S2686954322070207. EDN CVIVCS.

Статью рекомендовал к опубликованию д.т.н. Г.Е. Веселов.

**Мельников Сергей Юрьевич** – Российский университет дружбы народов имени Патриса Лумумбы; e-mail: melnikov@linfotech.ru; г. Москва, Россия; кафедра теории вероятностей и кибербезопасности института компьютерных наук и телекоммуникаций факультета физико-математических и естественных наук; д.ф.-м.н.; доцент.

**Мещеряков Роман Валерьевич** – ИПУ РАН; e-mail: mrv@ieee.org; г. Москва, Россия; г.н.с.; д.т.н.; профессор.

**Пересыпкин Владимир Анатольевич** – Академия криптографии РФ; e-mail: info@cryptoacademy.gov.ru; г. Москва, Россия; действительный член; д.т.н.

**Melnikov Sergey Yur'evich** – Patrice Lumumba Peoples' Friendship University of Russia; e-mail: melnikov@linfotech.ru; Moscow, Russia; the Department of Probability Theory and Cybersecurity, Institute of Computer Science and Telecommunications, Faculty of Physics, Mathematics and Natural Sciences; dr. of phys. and math. sc.; associate professor.

**Meshcheryakov Roman Valer'evich** – IPU RAS; e-mail: mrv@ieee.org; Moscow, Russia; chief researcher; dr. of eng. sc.; professor.

**Peresypkin Vladimir Anatol'evich** – Academy of Cryptography of the Russian Federation; e-mail: info@cryptoacademy.gov.ru; Moscow, Russia; Full Member; dr. of eng. sc.

УДК 004.382.2 DOI 10.18522/2311-3103-2024-5-68-78

**Е.А. Титенко, Э.И. Ватутин, М.А. Титенко, Э.В. Мельник, А.П. Локтионов**

# АППАРАТНО-ОРИЕНТИРОВАННЫЙ МЕТОД УСКОРЕННОГО ПОИСКА ВХОЖДЕНИЙ ОБРАЗЦА НА ОСНОВЕ СТРУКТУРНО-ПРОЦЕДУРНЫХ ВЫЧИСЛЕНИЙ

*Операция поиска вхождений образца в тексте является общезначимой в современных вычислительных средствах при решении проблемно-поисковых задач. Наибольший интерес представляют аппаратно-программные решения, имеющие однородную структуру и регулярные связи между вычислительными блоками. Целью работы является сокращение временных затрат на поиск вхождений на основе применения параллельного поиска в ассоциативной памяти и метода распараллеливания по итерациям. Предлагаемый метод использует ассоциативную память для параллельного поиска вхождений и динамическую реконфигурации структуры исходной строки из одномерного вида в матричную форму. Вовлечение в реконфигурацию всех элементов влечет избыточные затраты внутренней блочной памяти на последовательный просмотр частичных вхождений по одному множеству стартовых позиций, кратных длине образца (второй символьный операнд. Вместо этого предложен метод совмещения во времени поиска частичных вхождений по двум наборам подстрок, кратных длине образца, с одновременным пропорциональным уменьшением элементов разрядного среза ассоциативной памяти по каждому набору, что позволяет на текущем шаге поиска обрабатывать несколько символов образца. Количественные оценки времени поиска определяются количеством операций сравнения и записи подстрок в общем цикле работы, а также пропорциями времени данных операций. Показано, что для образцов более 10 элементов временной выигрыш составляет примерно в 1,8-2 раза. Данный эффект получен за счет исключения шагов последовательного сдвига с переходами между граничными элементами строк. Разработанный метод обеспечивает конвейерную обработку потока строковых операндов с совмещением просмотра на текущем шаге поиска неединичного множества символов обрабатываемой строки. Сокращение времени поиска обеспечивается введением конвейера, количество ступеней*

*которого зависит от коэффициента редукции размера разрядного среза, что позволяет аппаратно реализовать структурно-процедурный подход, применяемый в реконфигурируемых вычислительных системах.*

*Параллельный поиск; ассоциативная память; реконфигурация; метод итераций; конвейер.*

**E.A. Titenko, E.I. Vatutin, M.A. Titenko, E.V. Melnik, A.P. Loktionov**

## A HARDWARE-ORIENTED METHOD OF ACCELERATED SEARCH BY TEMPLATE BASED ON STRUCTURAL-PROCEDURAL COMPUTING

*The operation of searching for occurrences of a pattern in a text is generally significant in modern computing tools for solving problem-searching tasks. Of greatest interest are hardware and software solutions that have a homogeneous structure and regular connections between computing blocks. The aim of the work is to reduce the time costs for searching for occurrences based on the use of parallel search in associative memory and the method of parallelization by iterations. The proposed method uses associative memory for parallel search for occurrences and dynamic reconfiguration of the structure of the original string from a one-dimensional form to a matrix form. The method is critical to such resources as the number of memory access channels, the volume of block memory for creating and parallel operation of an array of associative cells. Involvement of all elements in the reconfiguration entails excessive costs of the internal block memory for sequential viewing of partial entries by one set of starting positions multiples of the sample length (the second symbolic operand). Instead, an approach is proposed to combine in time the search for partial entries by two sets of substrings multiples of the sample length, with a simultaneous proportional reduction in the elements of the bit slice of the associative memory for each set, which allows processing several sample symbols at the current search step. Quantitative estimates of search time are determined by the number of comparison and substring writing operations in the overall work cycle, as well as the proportions of the time of these operations. It is shown that for samples of more than 10 elements, the time gain is approximately 1.8-2 times. This effect is obtained by eliminating the steps of sequential shift with transitions between the boundary elements of the strings. The developed method provides pipeline processing of a stream of string operands with a combination of viewing at the current search step of a non-unit set of characters of the processed string. The search time is re duced by introducing a pipeline, the number of stages of which depends on the reduction coefficient of the bit slice size, which allows hardware implementation of the structural-procedural approach used in reconfigurable computing systems.*

*Parallel search; associative memory; reconfiguration; iteration method; pipeline.*

**Введение.** Научно-технический прогресс компьютерной индустрии направлен на повышение эффективности ее использования для различных масштабных задач управления, моделирования, поиска решений как на основе методов оптимизации, так и на основе методов вывода с применением баз знаний. Интенсификация вычислений в высокопроизводительных вычислительных системах (ВС) реализуется по нескольким направлениям: реконфигурируемые вычислительные системы, многопроцессорные кластерные системы, GRID-системы и др. Предметной областью являются вычислительно трудоемкие задачи проблемно-поискового и расчетного характера и аппаратно-программные средства их реализации или поддержки [1–6].

В первом случае интеллектуализация вычислений направлена на применение моделей и средств обработки знаний. Такие ВС призваны дополнить когнитивные способности человека в части систематизации информации, выделения новых знаний, поиску скрытых закономерностей и др. Они основаны на построении архитектур ВС, поддерживающих или использующих модели представления знаний и методы их обработки [7–10].

Второй путь интенсификации заключается в создании проблемно-ориентированных вычислительных средств, имеющих операционную часть, максимально приближенную к решаемой задаче. В настоящее время в России и за рубежом динамично развиваются реконфигурируемые вычислительные системы (РВС), построенные из программируемых логических интегральных схем (ПЛИС). Главным преимуществом РВС перед МВС с «жесткой» архитектурой является возможность адаптации РВС под структуру информационного графа задачи, что обеспечивает параллельно-конвейерную организацию вычислений без избыточных затрат на промежуточные пересылки и вычисления.

Использование ПЛИС в качестве элементной базы РВС имеет обоснованные перспективы. Объединение множества ПЛИС в единое вычислительное поле позволяет добиться практически линейного роста реальной производительности при масштабировании задач [11, 12]. Структурно-процедурный подход создания РВС на элементной базе ПЛИС заключается в непосредственном отображении топологии и структуры информационного графа задачи на вычислительную структуру (операционную часть) и в аппаратной поддержке базового подграфа в зависимости от имеющего ресурса ПЛИС (объем логических ячеек, размер блочной памяти, каналы доступа к памяти и др.).

Наиболее важной и трудоемкой операцией в слабо формализованных задачах является операция поиска по образцу (поиска вхождения образца в текст). Она лежит в основе работы абстрактных машин вывода, динамических экспертных систем, вычислительных машин, управляемых потоком данных, специализированных процессоров логического вывода, интеллектуальных аппаратных планировщиках и др. [13, 14].

Применение структурно-процедурного принципа вычислений для задачи поиска вхождений (поиска по образцу) позволяет получить ее базовый подграф и провести его преобразование к параллельно-конвейерной форме и аппаратно реализовать в ПЛИС, что делает общезначимыми метод и реализующее его устройство для высокоскоростных ВС [2, 10].

**Постановка задачи.** Известен огромный арсенал последовательных реализаций метода поиска по образцу, осуществляющих полную или частичную посимвольную обработку символьных операндов [15–17]. Тем не менее, одномерный формат операндов предполагает организацию последовательных процессов поиска, реализуя перемещения (переходы, прыжки, пропуски) через незначащих позиции в обрабатываемой строке. Такой подход имеет ряд недостатков при аппаратной реализации:

- зависимость времени поиска от длины обрабатываемой строки;
- невозможность досрочно прекратить поиск на текущем шаге;
- игнорирование возможности конвейеризации поиска.

Пусть в общем алфавите $A=\{a_1,a_2, \ldots a_d\}$ заданы конструктивные объекты:

- строка-образец $O$ длиной $n$ символов;
- исходный текст (обрабатываемая строка) $S$ длиной $m$ символов ($n \le m$).

Требуется найти позиции всех вхождений $O$ в $S$, т.е. разработать метод и алгоритм, находящие такие $i$, что

$$\forall i(O(1,n) = S(i,i+n-1)) \mid 1 \le i \le k, k = m-n+1, \quad (1)$$

где n>0, m>0 и n≤m.

Ведущими специалистами в современной информатике (Смит, Д. Гасфильд, Р. Бойер, Дж. Мур, и др. [15–18]) признано, что оценка времени работы поискового алгоритма осуществляется в количестве операций сравнения двух строк: строки-образца $O$ и текста $S$ или его текущей подстроки. Другими словами время поиска – это функция $T(n, m)$.

Методы поиска по образцу по применению аппаратных средств можно разбить на четыре класса [1–4]:

- алгоритмы поиска на микропроцессорах CISC-архитектурой;
- алгоритмы поиска на микропроцессорах RISC-архитектурой;
- алгоритмы поиска на графических микропроцессорах;
- алгоритмы поиска на элементной базе ПЛИС.

Первый класс алгоритмов поиска составляют алгоритмы, для которых повышение эффективности данных решений достигается за счет многоядерности и многопоточности для распараллеливания потока команд и управления переходами. Дополнительная информация о структурных отношениях закладывается в таблицы переходов на этапе подготовки данных.

Второй класс алгоритмов поиска составляют алгоритмы, для которых уменьшение количества сравнений в данных алгоритмах достигается за счет аппаратной поддержки набора строковых команд и форматов, позволяющих непосредственно вести обработку фрагментов подстрок с учетом структурных зависимостей. Построение и поддержка префиксов, суффиксов для строковых операндов позволяет сократить затраты времени на повторные шаги.

Третий класс построен на основе принципов адаптивного поиска и заключается в решении задачи приближенного сравнения двух строк с использованием методов обучения (систематизациями). Тем не менее, данный подход требует создания представительных обучающих и контрольных строковых выборок, имеющих значимые структурно-лингвистические характеристики, что позволяет точно установить отсутствие вхождения строки-образца в исходный текст за счет вычисления расстояния между эталоном и поисковым объектом. В случае ненулевого расстояния между ними данные алгоритмы не позволяют установить позицию вхождения образца, что ограничивает область эффективного применения адаптивного поиска.

Четвертый класс алгоритмов поиска (Shift-And, Сандер-Ву, Манбера, потоковый алгоритм поиска и др.) позволяет решать задачу поиска вхождений, непосредственно программируя информационный граф в вычислительную структуру на основе типовых ресурсов ПЛИС [6, 20].

В целом, анализ известных аппаратных решений поиска вхождений показывает, что все они имеют условия для распараллеливания вычислений, но этот параллелизм является неявным.

Вместе с тем, если размеры строковых операндов превышают предельные пороги ($n>10^2$ и $m>10^6$), то время работы алгоритмов становится практически неприемлемым [4] из-за лавинообразного роста количества промежуточных вариантов частичных вхождений строковых операндов своим фрагментами. Это означает, необходимы методы и алгоритмы поиска вхождений, не имеющие прямо пропорциональной временной зависимости от размеров операндов и ориентированные на передовые принципы интеллектуальной обработки информации («однородность-параллелизм», «ассоциация-контекст», «поток данных-конвейеризация», «реконфигурация-переменная структура»).

**Метод ускоренного ассоциативного поиска.** Ассоциативная память и основанные на ней ассоциативные процессоры параллельных вычислений [9, 21, 22] являются одним из значимых инструментов повышения эффективности вычислительных систем.

Как известно, ассоциативная память в качестве идентификатора ячеек использует не адрес, а контекст (контекстный атрибут), связанный (ассоциируемый) с данными, хранящимися в ячейках памяти [9] В силу наличия контекстной связи между идентификатором и данными разные ячейки памяти могут иметь одинаковый поисковый атрибут-контекст. Это обстоятельство стало основой поиска не ячейкам памяти, а по разрядным срезам.

Информационный граф задачи ассоциативного поиска показан на рис. 1, где $Я_{ij}$ – элементарная ячейка сравнения j-го бита контекстного атрибута $A_j$ и $ij$-го бита накопителя памяти$_j$, $M1_j$ – j-ый бит маски разрядных срезов, $M2_j$ – i-ый бит маски ячеек накопителя памяти, РгОтв – регистр ответов, АРБ – арбитр, $I_{вх}$, $I_{вых}$ – входные и выходные шины данных.

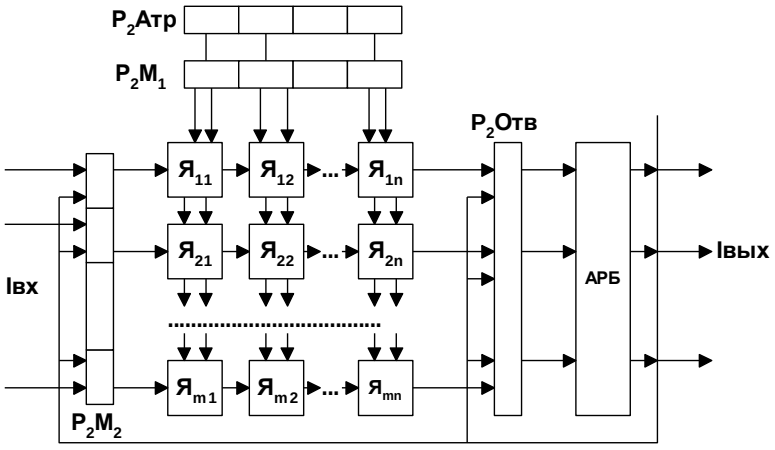



*Рис. 1. Модель ассоциативной памяти*

Основу известного метода ассоциативного поиска вхождений [23] составляет динамическая (циклическая) реконфигурация структуры исходной строки S: одномерный сдвиговый регистр ↔двумерная матрица. На четных шагах реконфигурации структуры *S* выполняется левый сдвиг на один символ по всей длине строки *S*, чем достигается совмещение во времени и пространстве всех подстрок из *S* длиной в *n* символов. На нечетных шагах реконфигурации *S* выполняется ассоциативный поиск (поиск по разрядным срезам) строки-образца по выделенным фрагментам из *S*. Двумерный вид *S* необходим для организации параллельного поиска всех вхождений по разрядным срезам ассоциативной запоминающей матрицы. Циклы поиска и реконфигурации соответственно выполняются *n-1* раз, что обеспечивает линейную временную сложность решения задачи *T – O(n)*.

Процесс ассоциативного поиска всех вхождений реализуется путем представления исходной строки *S* длиной до $m=n\times p$ символов в виде двухмерной матрицы из *p* строк по *n* разрядов в каждой строке, где *n* определяется разрядностью образца *O*. При таком представлении строки *S* образец *O* длиной в *n* символов параллельно сравнивается по *p* строкам матрицы в полном соответствии с известными способами ассоциативного поиска [22, 23].

Пусть для $A=\{0,1\}$ заданы $S=1100101011001001$ и $O=1001$ с длинами 16 элементов и 4 элемента соответственно. Процессы ассоциативного поиска вхождений по шагам отражены на рис. 2 (поиск, сдвиг, поиск). В случае положительного сравнения результат содержит логические «1» в строках ассоциативной матрицы с совпавшими фрагментами из *S*.

В случае отрицательного сравнения фиксируется нулевой результат, выполняется переход вперед на 1 символ (сдвиг влево на 1 символ строки S). В итоге выполняется переход к новой матрице с новым составом *p* подстрок. Цикл сравнения повторяется.

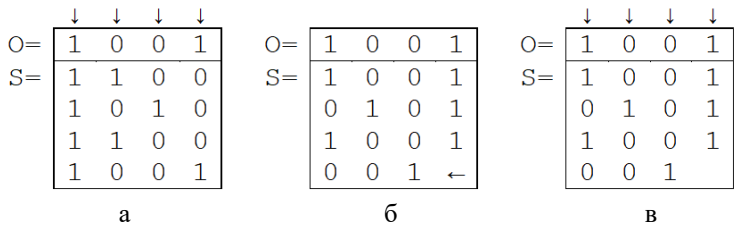



*Рис. 2. Шаги ассоциативного поиска с реконфигурацией S*

Информационный граф задачи тесно связан с моделью работы ассоциативной памяти (рис. 1) и поясняется на рис. 3, где показана параллельная обработка элементов каждого разрядного среза накопителя ассоциативной памяти и последовательная обработка разрядных срезов между собой. Основные вычислительные элементы в базовом подграфе α – двухвходовые компараторы на совпадение элемента $o_j$ строки-образца O и элемента $s_{ij}$ обрабатываемой строки (текста) S ($j=1 \ldots n$), $i=1 \ldots p$).

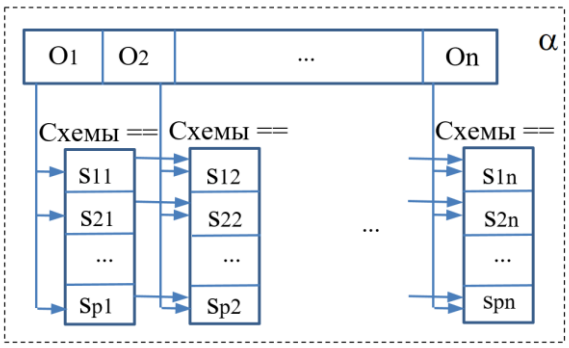



*Рис. 3. Информационный граф задачи ассоциативного поиска вхождений*

Двумерное представление обрабатываемой строки *S* в виде матрицы, позволяет вести параллельную обработку *p* строк в составе матрицы, принимая ширину матрицы равной длине образца, т.е. *n* элементов. Тогда переход к новой конфигурации поиска реализуется преобразованием к одномерному виду и сдвигу влево, чтобы получить в матрице новые *p* подстрок. Тем не менее, такой подход на текущем шаге не позволяет «смотреть вперед» и использовать дополнительную информацию для прекращения поиска по неперспективным строкам матрицы.

Недостаток разработанного метода ассоциативного поиска вхождений заключается в последовательной обработке разрядных срезов и невозможности организовать конвейерные вычисления по *n* разрядным срезам. Кроме того, переменный размер обрабатываемого строки S подразумевает использование неограниченного ресурса блочной памяти ПЛИС для поддержания матрицы $m=n\times p$ символов, при этом *p* – количество строк матрицы определяется размером задачи, что в общем случае приводит к последовательным решениям.

В методах структурно-процедурных вычислений, реализуемых в РВС, активно используется метод распараллеливания по итерациям [23]. Суть данного метода, заключается в том, что повторяющаяся последовательность операций, соответствующая базовому подграфу в вычислительной структуре, реализуется как конвейер из функциональныхе блоков с последовательной передачей данных между блоками и выполнении операций над разными частями матрицы. В результате чего при реализации в вычислительной структуре *n* однотипных блоков, соответствующих базовому подграфу, на выходе будет получен результат на *n*-й итерации. Итерационная зависимость обрабатываемых матриц позволяет её конвейеризовать для эффективной потоковой обработки. При этом базовый подграф формируется таким образом, чтобы сократить число обращений к распределенной памяти. Это прием позволяет сократить количество каналов доступа к памяти распараллелить обработку данных на текущей итерации.

Предлагаемый метод ускоренного ассоциативного поиска основан на следующих принципах:

- редукция количества элементов в разрядном срезе;
- совмещение в составе разрядного среза части элементов от текущего и следующего шагов поиска;
- замена последовательного левого сдвига строк матрицы на построчную перезапись;
- конвейеризация базового подграфа.

Сущность метода заключается в следующем. Однородная структура ассоциативной памяти позволяет объединить в пределах разрядного среза данные с *i*-ой и *i*+1-й, *i*+2-й, … позиций. Это обеспечивает на текущем шаге возможность анализа следующих частичных вхождений образца и использование этой информации для принятия решения о продолжении поиска. При объединении в текущем разрядном срезе ассоциативной памяти частей столбцов *V* матричного представления *S* возникает эффект продвижения вперед и учета большего количества частичных вхождений без реконфигурации матрицы. Продвижение вперед по подстрокам S согласуется с коэффициентом редукции *r* вектора V.

Далее рассматривается метод ускоренного ассоциативного поиска с объединением подстрок из S по 2 символа, т.е. реализуется «просмотр вперед» на 1 символ. Исходя из минимизации вычислительного ресурса принимается коэффициент редукции *r*=2.

Исходная строка S заранее разбивается на 2 части и представляется в виде

$$S = \{S\}^{1/2} \cup \{S\}^{1+/2}, \qquad (2)$$

где $\{S\}^{1/2}$ –половина строки с текущей позиции *i,*

$\{S^{1+/2}$ – половина строки, смещенная на 1 символ вперед.

Ниже показана разметка строки *S* для выделения и обработки двух смещенных подстрок $\{S\}^{1/2}$ и $\{S\}^{1+/2}$.

$$\{S\}^{1/2} = \underbrace{1100}_{11}\underbrace{1010}_{12}\underbrace{1100}_{13}\underbrace{0101}_{14}$$

$$\{S\}^{1+/2} = 1\overbrace{1001}^{21}\overbrace{0101}^{22}\overbrace{1000}^{23}\overbrace{0101}^{24}. \quad (3)$$

Шаг метода ускоренного ассоциативного поиска иллюстрируется на рис. 4.

| | | | | | |
|---|---|---|---|---|---|
| O= | 1 | 0 | 0 | 1 | |
| S= | | | | | |
| {S}1/2 | 1 | 1 | 0 | 0 | 0 V1i |
| | 1 | 0 | 1 | 0 | 0 |
| {S}1+/2 | 1 | 0 | 0 | 1 | 1 V2i |
| | 0 | 1 | 0 | 1 | 0 |

*Рис. 4. Ускоренный ассоциативный поиск с просмотром вперед на 1 символ*

При такой организации в каждом разрядном срезе верхняя половина элементов записана от текущей позиции поиска по S, нижняя половина – записана от следующей позиции поиска по строке S. В результате ассоциативный поиск выполняется по 2 матрицам одновременно, что позволяет принимать обоснованное решение о следующем шаге поиска (при отрицательных результатах по векторам ответов $V1$ и $V2$).

Далее цикл поиска с просмотром «вперед» далее выполняется над вторыми половинами матрицы $S$, также одна из которых будет смещена на 1 символ вперед.

В соответствии со структурно-процедурными принципами вычислений выполняется распараллеливание по итерациям. В результате получается двухконвейерная вычислительная структура (рис. 5), совмещающая во времени потоковую обработку подготовленных строковых операндов: строки-образца P и подстрок {S} матрицы.

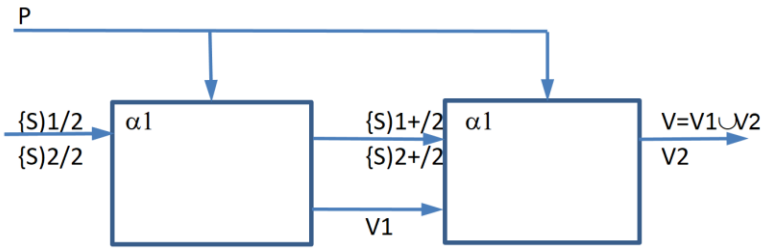



*Рис. 5. Двухступенчатый конвейер поиска вхождений*

**Результаты и обсуждение.** Аппаратная реализация поисковых операций в ассоциативной памяти (поиск на совпадение, несовпадение и приблизительный поиск) и организация конвейерной обработки матриц позволяют получить временной выигрыш при использовании метода ускоренного ассоциативного поиска. Данный аппаратно-ориентированный метод отличают:

- увеличенный на коэффициент редукции $r$ объем вычислительного ресурса для создания ступеней конвейера;
- сохраненное число каналов доступа к памяти;
- возможность параллельной перезаписи по строкам нижней половины матрицы в ее верхнюю часть без обращения к внешней памяти;
- сокращенный размер подгружаемой подстроки.

Известный метод ассоциативного поиска с динамической реконфигурацией обрабатываемой строки $S$ имеет оценку временной сложности

$$T(n)=(n\text{-}1)t_{COMP}+(n\text{-}1)t_{SHIFT}. \quad (4)$$

Разработанный метод использует динамическую реконфигурацию части обрабатываемой строки только для первого и последнего символов образца, заменяя последовательные сдвиги на загрузку фрагмента строки в ячейки ассоциативной памяти. В результате оценка временной сложности по методу ускоренного поиска составляет

$$T(n)=n/2t_{COMP}+(n/2)t_{SAVE}+2t_{SHIFT}. \quad (5)$$

Анализ времени базовых операций при ассоциативном поиске (компарирование, сдвиг, загрузка строки) показывает, что они имеют следующие соотношения: $t_{COMP}=\tau$, $t_{SHIFT}=0{,}6\tau$, $t_{SAVE}=0{,}4\tau$.

Тогда временные затраты на параллельный поиск вхождений образца на основе ассоциативной памяти по (4) и (5) имеют вид соответственно

$$T(n)=(n\text{-}1)\cdot 1{,}8\,\tau+C_1, \quad (6)$$

$$T(n)=0{,}7n\tau+1{,}6\,\tau+C_2. \quad (7)$$

На рис. 6 показаны временные зависимости разработанного метода ускоренного поиска (Метод 2 ) и известного метода ассоциативного поиска (Метод 1) для переменной длины образца $n$=4, 8, 12, 16, 20, 24, 28, 32 элементов, приняв следующие временные задержки (в условных единицах времени) $\tau=15$, $C_1=24$, $C_1=36$.

Анализ графиков (рис. 6) показывает, что разработанный метод имеет эффективную область применения при $n \geq 10$ элементов. Малые размеры длины образца $3<n<8$ делают предпочтительным поиск вхождений образца на основе динамической реконфигурации «строка ↔матрица».

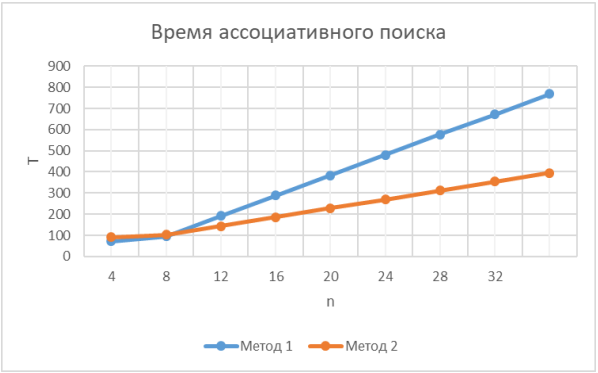



*Рис. 6. Временные зависимости поиска вхождений*

Распараллеливание по итерациям для метода ускоренного поиска при $n<8$ по-прежнему предполагает проведение 2 шагов динамической реконфигурации для первого и последнего символов строки $S$, что негативно сказывается на времени поиска. При этом увеличенные вычислительные ресурсы на организацию двухступенчатого конвейера не компенсируются временным выигрышем в силу малого числа итераций.

**Выводы.** На основе метода ассоциативного поиска разработан метод ускоренного поиска вхождений образца, отличающийся применением принципов структурно-процедурных вычислений. Отображение информационного графа ассоциативного поиска в матричную вычислительную структуру, состоящую из массива параллельно работающих символьных компараторов, позволяет совместить поиск по смежным символам образца путем создания конвейера. Применение принципа распараллеливания по итерациям позволяет сохранить неизменным критичный вычислительный ресурс при обработке матриц – число каналов доступа к памяти. Получен временной выигрыш, составивший примерно 2 раза по сравнению с ассоциативным поиском вхождений. Разработанный метод эффективно применим для образцов с длинами от 10 элементов, что преимущественно встречается в машинах баз знаний, динамических экспертных системах, вычислительных средствах поддержки трейдинговых процессов, анализа Big Data и др. [24–26].

БИБЛИОГРАФИЧЕСКИЙ СПИСОК


1. *Воеводин В.В.* Математические модели и методы в параллельных процессах. – М.: Наука, 1986. – 296 с.
2. *Гузик В.Ф., Каляева И.А., Левин И.И.* Реконфигурируемые вычислительные системы: учеб. пособие. – Таганрог: Изд-во ЮФУ, 2016. – 472 с.
3. *Корнеев В.В.* Вычислительные системы. – М.: Гелиос АРВ, 2004. – 510 с.
4. *Бурцев В.С.* Параллелизм вычислительных процессов и развитие архитектуры суперЭВМ: Сб. статей / сост. В.П. Торчигин, Ю.Н. Никольская, Ю.В. Никитин. – М.: ТОРУС ПРЕСС, 2006. – 416 с.
5. *Левин И.И., Федоров А.М., Доронченко Ю.И., Раскладкин М.К.* Перспективные высокопроизводительные реконфигурируемые вычислители с иммерсионным охлаждением // Известия ЮФУ. Технические науки. – 2020. – № 7 (217). – С. 6-19. – DOI: 10.18522/2311-3103-2020-7-6-19.
6. *Каляев А.В., Левин И.И.* Модульно-наращиваемые многопроцессорные системы со структурно-процедурной организацией вычислений. – М.: Изд-во "Янус-К", 2003. – 380 с.
7. *Lothaire M.* Applied Combinatorics on Words. In: Encyclopedia of Mathematics and its Applications. – Cambridge: Cambridge University Press, 2005.
8. *Люгер Дж.Ф.* Искусственный интеллект: стратегии и методы решения сложных проблем. – М.: Издательский дом «Вильямс». 2003. – 864 с.
9. *Огнев И.В., Борисов В.В., Сутула Н.А*. Ассоциативные память, среды, системы. – М.: Горячая линия – Телеком, 2016. – 420 с.
10. *Адамов А.А., Эйсымонт Л.К.* Варианты архитектурных решений ЭКБ для систем искусственного интеллекта // Проектирование будущего. Проблемы цифровой реальности: Тр. 3-й Международной конференции. – М.: ИПМ им. М.В. Келдыша, 2020. – С. 112-131.
11. *Левин И.И., Пелипец А.В.* Методология распараллеливания по итерациям при решении задач линейной алгебры на реконфигурируемых вычислительных системах // Вестник компьютерных и информационных технологий. – 2016. – № 7 (145). – С. 34-40.
12. *Левин И.И., Подопригора А.В.* Модифицированный метод обработки больших разреженных неструктурированных матриц на реконфигурируемых вычислительных системах // Вычислительные методы и программирование. – 2024. – Т. 25, № 2. – С. 142-154.
13. *Добрица В.П., Титенко Е.А., Халин Ю.А., Киселев А.В*. Системы искусственного интеллекта. – Курск: ЗАО «Университетская книга», 2023. – 143 с.
14. *Эйсымонт Л.К., Моляков А.С., Заборовский В.С., Федоров С.А.* Символьная обработка: эпизоды отечественной истории и перспективы // Матер. 2-й Всероссийской научно-технической конференции «Суперкомпьютерные технологии (СКТ-2012)», с. Дивноморское, 2012. – С. 202-206.
15. *Макконелл Дж.* Основы современных алгоритмов. – 2-е изд., доп. – М.: Техносфера, 2004. – 368 с.
16. *Окулов С.М.* Алгоритмы обработки строк. – М.: БИНОМ. Лаборатория знаний, 2009. – 255 с.
17. *Рыбина Г.В.* Основы построения интеллектуальных систем. – М.: Финансы и статистика, 2010. – 430 с.
18. *Hume A., Sunday D*. Fast string searching // Software Practice and Experience. – November 1991. – Vol. 21, No. 11. – P. 1221-48.
19. *Попов Э.В.* Статические и динамические экспертные системы. – М.: Финансы и статистика, 1996. – 211 с.
20. *Титенко Е.А., Титов В.С., Коновальчик А.П.* Высокопроизводительные вычислительные системы на основе ПЛИС // Известия Юго-Западного государственного университета. – 2012. – № 4-2 (43). – С. 73.
21. *Емельянов С.Г., Титенко Е.А., Зерин И.С.* Однородные вычислительные структуры для параллельных символьных вычислений // Известия Юго-Западного государственного университета. – 2011. – № 6-2 (39). – С. 77а-82.
22. *Типикин А.П., Титенко Е.А.* Модификация цикла работы машины вывода для параллельных вычислительных устройств // Известия Юго-Западного государственного университета. – 2011. – № 6-2 (39). – С. 92-96.
23. *Зерин И.С., Атакищев О.И., Титенко Е.А. [и др.].* Метод, алгоритм и техническое решение параллельного поиска и подстановки на ассоциативной памяти // В мире научных открытий. – 2012. – № 1-1 (25).
24. *Левин И.И., Подопригора А.В.* Метод распараллеливания по базовым макрооперациям для обработки больших разреженных неструктурированных матриц на РВС // Известия ЮФУ. Технические науки. – 2022. – № 6 (230). – С. 72-83.

25. *Добрица В.П., Титенко Е.А., Халин Ю.А., Катыхин А.И.* Модели представления и обработки знаний в информационно-аналитических системах. – Курск: ЗАО «Университетская книга», 2023. – 172 с.
26. *Овчинкин О.В., Титова Г.С., Халин Ю.А. [и др.]*. Исследование систем управления: учеб. пособие. – Курск: ЗАО Университетская книга, 2018. – 172 с.

REFERENCES

1. *Voevodin V.V.* Matematicheskie modeli i metody v parallel'nykh protsessakh [Mathematical models and methods in parallel processes]. Moscow: Nauka, 1986, 296 p.
2. *Guzik V.F., Kalyaea I.A., Levin I.I.* Rekonfiguriruemye vychislitel'nye sistemy: ucheb. posobie [Reconfigurable computing systems: textbook]. Taganrog: Izd-vo YuFU, 2016, 472 p.
3. *Korneev V.V.* Vychislitel'nye sistemy [Computing systems]. Moscow: Gelios ARV, 2004, 510 p.
4. *Burtsev V.S.* Parallelizm vychislitel'nykh protsessov i razvitie arkhitektury superEVM: Sb. statey [Parallelism of computing processes and development of supercomputer architectures: Collection of articles], compilers V.P. Torchigin, Yu.N. Nikol'skaya, Yu.V. Nikitin. Moscow: TORUS PRESS, 2006, 416 p.
5. *Levin I.I., Fedorov A.M., Doronchenko Yu.I., Raskladkin M.K.* Perspektivnye vysokoproizvoditel'nye rekonfiguriruemye vychisliteli s immersionnym okhlazhdeniem [Promising high-performance reconfigurable computers with immersion cooling], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2020, No. 7 (217), pp. 6-19. DOI: 10.18522/2311-3103-2020-7-6-19.
6. *Kalyaev A.V., Levin I.I.* Modul'no-narashchivaemye mnogoprotsessornye sistemy so strukturno-protsedurnoy organizatsiey vychisleniy [Modular-scalable multiprocessor systems with structural-procedural organization of computations]. Moscow: Izd-vo "Yanus-K", 2003, 380 p.
7. *Lothaire M.* Applied Combinatorics on Words. In: Encyclopedia of Mathematics and its Applications. Cambridge: Cambridge University Press, 2005.
8. *Lyuger Dzh.F.* Iskusstvennyy intellekt: strategii i metody resheniya slozhnykh problem [Artificial intelligence: strategies and techniques for solving complex problems]. Moscow: Izdatel'skiy dom «Vil'yams». 2003, 864 p.
9. *Ognev I.V., Borisov V.V., Sutula N.A.* Assotsiativnye pamyat', sredy, sistemy [Associative memory, environments, systems]. Moscow: Goryachaya liniya – Telekom, 2016, 420 p.
10. *Adamov A.A., Eysymont L.K.* Varianty arkhitekturnykh resheniy EKB dlya sistem iskusstvennogo intellekta [Variants of architectural solutions for electronic components for artificial intelligence systems], *Proektirovanie budushchego. Problemy tsifrovoy real'nosti: Tr. 3-y Mezhdunarodnoy konferentsii* [Designing the future. Problems of digital reality: proceedings of the 3rd International Conference.]. Moscow: IPM im. M.V. Keldysha, 2020, pp. 112-131.
11. *Levin I.I., Pelipets A.V.* Metodologiya rasparallelivaniya po iteratsiyam pri reshenii zadach lineynoy algebry na rekonfiguriruemykh vychislitel'nykh sistemakh [Methodology of parallelization by iterations in solving problems of linear algebra on reconfigurable computing systems], *Vestnik komp'yuternykh i informatsionnykh tekhnologiy* [Bulletin of Computer and Information Technologies], 2016, No. 7 (145), pp. 34-40.
12. *Levin I.I., Podoprigora A.V.* Modifitsirovannyy metod obrabotki bol'shikh razrezhennykh nestrukturirovannykh matrits na rekonfiguriruemykh vychislitel'nykh sistemakh [Modified method for processing large sparse unstructured matrices on reconfigurable computing systems], *Vychislitel'nye metody i programmirovanie* [Computational methods and programming], 2024, Vol. 25, No. 2, pp. 142-154.
13. *Dobritsa V.P., Titenko E.A., Khalin Yu.A., Kiselev A.V.* Sistemy iskusstvennogo intellekta [Artificial intelligence systems]. Kursk: ZAO «Universitetskaya kniga», 2023, 143 p.
14. *Eysymont L.K., Molyakov A.S., Zaborovskiy V.S., Fedorov S.A.* Simvol'naya obrabotka: epizody otechestvennoy istorii i perspektivy [Symbolic processing: episodes of Russian history and prospects], *Mater. 2-y Vserossiyskoy nauchno-tekhnicheskoy konferentsii «Superkomp'yuternye tekhnologii (CKT-2012)», s. Divnomorskoe, 2012* [Materials of the 2nd All-Russian Scientific and Technical Conference “Supercomputer Technologies (SKT-2012)”, With. Divnomorskoe, 2012], pp. 202-206.
15. *Makkonell Dzh.* Osnovy sovremennykh algoritmov [Fundamentals of modern algorithms]. 2nd ed. Moscow: Tekhnosfera, 2004, 368 p.
16. *Okulov S.M.* Algoritmy obrabotki strok [String processing algorithms]. Moscow: BINOM. Laboratoriya znaniy, 2009, 255 p.
17. *Rybina G.V.* Osnovy postroeniya intellektual'nykh system [Fundamentals of building intelligent systems]. Moscow: Finansy i statistika, 2010, 430 p.
18. *Hume A., Sunday D*. Fast string searching, *Software Practice and Experience,* November 1991, Vol. 21, No. 11, pp. 1221-48.

19. *Popov E.V.* Staticheskie i dinamicheskie ekspertnye sistemy [Static and dynamic expert systems]. Moscow: Finansy i statistika, 1996, 211 p.
20. *Titenko E.A., Titov V.S., Konoval'chik A.P.* Vysokoproizvoditel'nye vychislitel'nye sistemy na osnove PLIS [High-performance computing systems based on FPGA], *Izvestiya Yugo-Zapadnogo gosudarstvennogo universiteta* [Bulletin of the South-West State University], 2012, No. 4-2 (43), pp. 73.
21. *Emel'yanov S.G., Titenko E.A., Zerin I.S.* Odnorodnye vychislitel'nye struktury dlya parallel'nykh simvol'nykh vychisleniy [Homogeneous computing structures for parallel symbolic computations], *Izvestiya Yugo-Zapadnogo gosudarstvennogo universiteta* [Bulletin of the South-West State University], 2011, No. 6-2 (39), pp. 77a-82.
22. *Tipikin A.P., Titenko E.A.* Modifikatsiya tsikla raboty mashiny vyvoda dlya parallel'nykh vychislitel'nykh ustroystv [Modification of the output machine operation cycle for parallel computing devices], *Izvestiya Yugo-Zapadnogo gosudarstvennogo universiteta* [Bulletin of the South-West State University], 2011, No. 6-2 (39), pp. 92-96.
23. *Zerin I.S., Atakishchev O.I., Titenko E.A. [i dr.].* Metod, algoritm i tekhnicheskoe reshenie parallel'nogo poiska i podstanovki na assotsiativnoy pamyati [Method, algorithm and technical solution for parallel search and substitution on associative memory], *V mire nauchnykh otkrytiy* [In the world of scientific discoveries], 2012, No. 1-1 (25).
24. *Levin I.I., Podoprigora A.V.* Metod rasparallelivaniya po bazovym makrooperatsiyam dlya obrabotki bol'shikh razrezhennykh nestrukturirovannykh matrits na RVS [Method of parallelization by basic macro-operations for processing large sparse unstructured matrices on RCS], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2022, No. 6 (230), pp. 72-83.
25. *Dobritsa V.P., Titenko E.A., Khalin Yu.A., Katykhin A.I.* Modeli predstavleniya i obrabotki znaniy v informatsionno-analiticheskikh sistemakh [Models of representation and processing of knowledge in information and analytical systems]. Kursk: ZAO «Universitetskaya kniga», 2023, 172 p.
26. *Ovchinkin O.V., Titova G.S., Khalin Yu.A. [i dr.].* Issledovanie sistem upravleniya: ucheb. posobie [Research of control systems: textbook]. Kursk: ZAO Universitetskaya kniga, 2018, 172 p.

Статью рекомендовал к опубликованию д.т.н., профессор А.В. Боженюк.

**Титенко Евгений Анатольевич** – Юго-Западный государственный университет; e-mail: johntit@mail.ru; г. Курск, Россия; тел.: +79051588904; кафедра программной инженерии; к.т.н.; доцент.

**Ватутин Эдуард Игоревич** – e-mail: evatutin@mail.ru; д.т.н.; доцент; профессор кафедры вычислительной техники.

**Титенко Михаил Андреевич** – e-mail: mikhail-titenko@mail.ru; аспирант.

**Локтионов Аскольд Петрович** – e-mail: loapa@mail.ru; д.т.н.; доцент; главный научный сотрудник центра по координации академической и вузовской науки, научной и образовательной деятельности.

**Мельник Эдуард Всеволодович** – Южный федеральный университет; e-mail: evm17@mail.ru; г. Таганрог, Россия; г.н.с.; зав. лабораторией интеллектуальных технологий.

**Titenko Evgeny Anatolievich** – South-West State University; e-mail: johntit@mail.ru; Kursk, Russia; phone: +79051588904; the Department of Software Engineering; cand. of eng. sc.; associate professor.

**Vatutin Eduard Igorevich** – e-mail: evatutin@mail.ru; dr. of eng.; sc.; associate professor; professor of the Computer Science Department.

**Titenko Mikhail Andreevich** – e-mail: mikhail-titenko@mail.ru; postgraduate student.

**Loktionov Askold Petrovich** – e-mail: loapa@mail.ru; dr. of eng. sc.; associate professor; chief researcher of the Center for Coordination of Academic and University Science, Scientific and Educational Activities.

**Melnik Eduard Vsevolodovich** – Southern Federal University; e-mail: evm17@mail.ru; Taganrog, Russia; chief researcher; head of the Laboratory of Intelligent Technologies.

УДК 004.056 DOI 10.18522/2311-3103-2024-5-79-88

**И.В. Машкина, А.М. Уразаева**

# МЕТОД РАЗРАБОТКИ БАЗЫ ЗНАНИЙ СЦЕНАРИЕВ УГРОЗ ДЛЯ СИСТЕМЫ РЕАГИРОВАНИЯ НА ИНЦИДЕНТЫ (IRP)

*Цель работы – исследование возможности повышения эффективности реагирования на инциденты информационной безопасности (ИБ). Это может быть достигнуто путем разработки системы, способной быстро локализовать инцидент, обеспечивающей автоматизацию реагирования на угрозу ИБ, предпринимающей заранее заданные действия в зависимости от деталей реализуемого сценария угрозы. Предложена архитектура построения IRP-системы, основными модулями которой являются база знаний сценариев реагирования, база знаний сценариев угроз, модули определения статуса инцидента и принятия решений по формированию командной информации. Решена задача разработки сценариев угроз для создания базы знаний сценариев, на основе которой могут быть разработаны адекватные сценарии реагирования, уникальные для каждой цепочки последовательности действий киберпреступника, событий и задействованных объектов. Формализован метод разработки базы знаний сценариев угроз на основе построения EPC-диаграмм сценариев, отображающих многокомпонентные атаки с учетом тактик, техник, используемых уязвимостей, угроз безопасности информации (УБИ), приведенных в нормативных документах и базах данных. Сформулированы правила построения EPC-диаграмм сценариев угроз и методика EPC-моделирования для объектов воздействия в АСУ ТП. Рассмотрен пример сценария атаки на промышленную сеть из глобальной сети в случае, когда киберпреступник, атаковав компьютер удаленного пользователя, в первую очередь осуществляет несанкционированный доступ в корпоративный сегмент, закрепляется в нем для дальнейшего проникновения за периметр технологической сети. Приведена разработанная EPC-диаграмма сценария угрозы с указанием используемых тактик, техник, промежуточных УБИ, некоторых уязвимостей. Формализована оценка вероятности реализации сценария.*

*Система реагирования на инциденты; тактики; техники; уязвимости; УБИ; база знаний сценариев угроз; база знаний сценариев реагирования; EPC-диаграмма.*

**I.V. Mashkina, A.M. Urazaeva**

# METHOD OF DEVELOPMENT OF THREAT SCENARIOS KNOWLEDGE BASE FOR INCIDENT RESPONSE PLATFORM (IRP)

*The objective of the work is to study the possibility of increasing the efficiency of response to information security (IS) incidents. This can be achieved by developing a system capable of quickly localizing an incident, providing automation of response to an IS threat, taking predetermined actions depending on the details of the threat scenario being implemented. An architecture for constructing an IRP system is proposed, the main modules of which are a response scenario knowledge base, a threat scenario knowledge base, modules for determining the incident status and making decisions on the formation of command information. The problem of developing threat scenarios for creating a scenario knowledge base has been solved, on the basis of which adequate response scenarios can be developed that are unique for each chain of the cybercriminal's actions, events and involved objects. The paper formalizes the method for developing a knowledge base of threat scenarios based on constructing EPC diagrams of scenarios that display multi-component attacks taking into account tactics, techniques, vulnerabilities used, and information security threats (IST) specified in regulatory documents and databases. The paper formulates the rules for constructing EPC diagrams of threat scenarios and the methodology for EPC modeling for objects of influence in ICS. An example of an attack scenario on an industrial network from a global network is considered in the case when a cybercriminal, having attacked a remote user's computer, first gains unauthorized access to the corporate segment and gains a foothold in it for further penetration beyond the perimeter of the process network. The paper presents the developed EPC diagram of a threat scenario indicating the tactics, techniques, intermediate IST, and some vulnerabilities used. The assessment of the probability of scenario implementation is formalized.*

*Incident response system; tactics; technics; vulnerabilities; information security threats; threat scenarios knowledge base; EPC -diagram.*

**Введение**. Приказы ФСТЭК России №31 и №239 утверждают требования к обеспечению защиты информации в АСУ ТП на критически важных и потенциально опасных объектах: атомной энергетики, добычи и транспортировки нефти и газа, оборотно-промышленного комплекса и транспорта [1, 2]. Специалисты в области ИБ отмечают, что в 2024 году проблемы защиты АСУ ТП остаются крайне актуальными и сложными [3–5]. Среди ключевых проблем, таких как: интеграция АСУ ТП с корпоративными ИТ-сетями, увеличение числа целенаправленных атак на системы управления технологическими процессами, необходимость обеспечения совместимости средств защиты информации с импортозамещающими отечественными SCADA-системами и ПЛК, – особо отмечаются нереализованные на многих объектах процессы мониторинга и реагирования на инциденты.

Отсутствие системы мониторинга, анализа угроз и реагирования на инциденты может привести к нарушению киберустойчивости промышленной системы и, следовательно, к нарушению непрерывности технологического процесса, увеличению времени на восстановление после атаки.

В последние годы особую актуальность приобрела тематика автоматизации реагирования на угрозы ИБ. Некоторые SIEM обладают встроенной IRP-системой, способной быстро локализовать инцидент и уменьшить или исключить разрушительность последствий.

Несколько самостоятельных IRP решений российского производителя для выполнения базовых задач находятся в промышленной эксплуатации: Jet Signal компании «Инфосистемы Джет», R-Vision компании «Р-Вижн», Security Vision компании «Интеллектуальная безопасность» [6–8]. На российском рынке представлены также продукты Израильской компании CyberBit SOC 3D, а также разработка компании IBM IBM Resilient IRP [9, 10].

**Архитектура IRP**. IRP (Incident Response Platform) – это система автоматизации реагирования на инциденты кибербезопасности, которая выполняет функции по сбору дополнительной информации, сдерживанию, устранению угрозы либо восстановлению системы после атаки, а также по структурированию данных о расследовании инцидента [11].

На рис. 1 приведена предлагаемая архитектура построения IRP системы. Основными модулями являются база знаний сценариев реагирования и база знаний сценариев угроз. Причем база сценариев реагирования может быть разработана на основе полного перечня всех возможных типов инцидентов ИБ, то есть на основе модели угроз конкретному объекту защиты, которая, как известно [12], включает в себя список актуальных угроз. Таким образом, эффективные сценарии реагирования на киберинциденты могут быть разработаны только с учетом сценариев угроз. На основе модели сценария угрозы формируется адекватный сценарий реагирования, уникальный для каждой последовательности событий и задействованных объектов. Сценарий реагирования представляет собой совокупность правил и выполняемых действий, специфичных для индикаторов – признаков этапов реализуемого сценария угрозы.

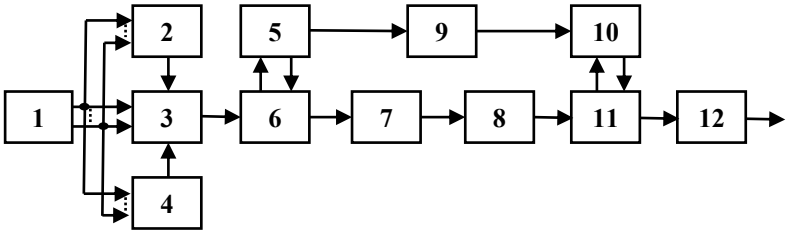



*Рис. 1. Архитектура системы реагирования на инциденты*

1 – источники данных о событиях безопасности (с SIEM системы);
2 – модуль определения задействованных объектов инфраструктуры;
3 – модуль анализа данных (техник, индикаторов и др.);
4 – модуль контроля привилегий;
5 – база знаний сценариев угроз целевым объектам инфраструктуры;

6 – модуль определения сценария угрозы в реальном времени;

7 – модуль численной оценки риска реализации угрозы;

8 – модуль определения статуса инцидента;

9 – модуль разработки плана реагирования;

10 – база знаний сценариев реагирования, адаптированных под конкретные сценарии угроз;

11 – модуль принятия решений;

12 – модуль формирования командной информации (Active Response) на агенты реагирования (запуск скриптов, воздействие на средства защиты и др.).

Рассмотрим, как может быть решена задача разработки сценариев угроз для создания базы знаний угроз и сценариев реагирования на их основе. В работах [13, 14] было предложено для моделирования угроз использовать принципы методологии ARIS [15]. Разработка EPC-диаграмм сценариев угроз позволяет отразить на одной схеме все значимые этапы многокомпонентной атаки. Четко обозначенные события, как результаты развития атаки, позволяют отметить и настроить точки контроля индикаторов для всех значимых этапов.

EPC-диаграмма процесса реализации угрозы целевому объекту АСУ ТП может быть представлена в виде комбинации событий и функций. При этом под функцией понимаем проводимые исполнителем-киберпреступником тактики и техники. Таким образом, функция – это действие или набор действий, выполняемых в информационной среде объекта защиты киберпреступником; функция может быть поименована соответствующей техникой из приложения 11 методики [12].

EPC-диаграмма сценария угрозы представляет собой отображение (сверху вниз) последовательности выполнения техник, начиная от исходного действия киберпреступника до достижения им цели – угрозы безопасности информации (УБИ) объекту воздействия, это может быть УБИ из банка данных [16]. Реализация на определенном этапе техники или их совокупности вызывает событие – состояние информационной среды, которое может быть оценено как некоторая промежуточная УБИ, существенная для достижения объекта воздействия угрозы. Эта промежуточная УБИ оказывает влияние на дальнейшее развитие сценария. Событие отображается на диаграмме как специальный элемент. В свою очередь событие – промежуточная УБИ – активизирует последующие тактики. Функции и события в процессе реализации сценария чередуются. Решение о развитии сценария, т.е. ходе выполнения многокомпонентной атаки, принимается киберпреступником по мере поиска уязвимостей для реализации техник.

**Правила построения EPC-диаграмм сценариев многокомпонентных атак** следующие:

- ♦ успешное выполнение киберпреступником техники приводит одновременно к нескольким событиям, для отображения на диаграмме используется оператор AND;
- ♦ промежуточная УБИ реализуется после одновременного выполнения двух техник, для отображения используется оператор AND перед УБИ;
- ♦ очередная техника может быть выполнена только после реализации двух УБИ, это отображается с помощью оператора AND перед техникой;
- ♦ промежуточная УБИ обеспечивает возможности выполнения двух техник, используется оператор AND перед техниками;
- ♦ выполнение техники может привести к реализации одной из двух или двух УБИ, на диаграмме это обозначается с помощью оператора OR после техники и перед $УБИ_i$ и $УБИ_j$;
- ♦ УБИ может быть реализована после выполнения одной из двух или двух техник, используется оператор OR перед УБИ;
- ♦ техника сможет быть выполнена после реализации одного из двух событий, или двух одновременно, используется оператор OR перед техникой;
- ♦ выполнение какой-либо техники может привести только к одному из событий: удалось реализовать УБИ или не удалось, это обозначается на диаграмме с помощью оператора XOR перед событиями;

♦ УБИ реализуется в результате наличия уязвимостей для выполнения только одной из двух техник, используется оператор XOR перед УБИ;

♦ техника может быть выполнена сразу после реализации одной из УБИ, тогда используется оператор XOR перед техникой.

**Методика EPC-моделирования сценариев многокомпонентных атак** заключается в следующем:

♦ задание целевого объекта воздействия угрозы в АСУ ТП: SCADA система, АРМ инженера-программиста по управлению процессом, OPC-сервер, АРМ оператора, PLC-контроллер, рабочая станция бизнес-контура сети с технологической сетью;

♦ определение источника угрозы: киберпреступник, удаленный пользователь-нарушитель, внутренний нарушитель;

♦ определение состояния информационной среды на начало реализации сценария;

♦ выявление возможной начальной тактики и соответствующих техник;

♦ обозначение событий, к которым могут привести реализуемые техники;

♦ оценка наличия уязвимостей информационной среды, эксплуатация которых позволит нарушителю выполнение техник в логической цепочке взаимодействия событий и функций;

♦ поиск промежуточных УБИ многокомпонентной атаки;

♦ достижение нарушителем цели реализации сценария – возможности нарушения свойства безопасности информации объекта воздействия угрозы;

♦ получение численной оценки вероятности реализации сценария угрозы.

**Построение сценария атаки на корпоративный сегмент промышленного предприятия с технологической сетью.** По статистике NIST [17] самыми опасными являются целенаправленные внешние атаки.

Для реализации атаки на промышленную сеть через глобальную, киберпреступник в первую очередь должен осуществить несанкционированный доступ в корпоративный сегмент и закрепиться в нем для дальнейшего проникновения за периметр промышленной сети.

Рассмотрим пример, когда удаленному пользователю предоставляется доступ в корпоративный сегмент (бизнес-контур) через VPN-соединение. В этом случае уровень защищенности сети компании, в том числе технологической сети, становится зависимым от защищенности компьютера удаленного пользователя. Если должные меры безопасности не реализованы или имеются уязвимости, киберпреступник, атаковав узел, может использовать VPN для туннелирования трафика за периметр корпоративного сегмента сети предприятия с АСУ ТП.

В результате заражения компьютера удаленного пользователя вредоносным программным обеспечением, оказываются скомпрометированы аутентификационные данные пользователя, который может иметь право доступа к каким-либо серверам бизнес-контура. Таким образом, с помощью троянской программы, которая собрала логины, пароли и другие данные, необходимые для связи с сервером бизнес-контура по каналу VPN, киберпреступник создает на своем компьютере копию скомпрометированного узла и осуществляет попытку нарушения периметра и далее подключения к серверу бизнес-контура сети, имея данные идентификации и аутентификации удаленного легитимного пользователя.

При наличии ошибок настройки прав доступа пользователей к ресурсам сервера киберпреступник после авторизации на сервере может определить директорию, в которой хранятся хэш-суммы паролей пользователей. Далее возможна процедура обратного пересчета хэш-функций, перебор собранных паролей и попытка авторизоваться на сервере под логином администратора. Права администратора позволяют ему, изменив конфигурацию сетевого оборудования, получить доступ в АСУ ТП для реализации атаки на целевой объект воздействия. На рис. 2 приведена разработанная EPC-диаграмма сценария атаки на корпоративный сегмент промышленного предприятия. В табл. 1 приведен перечень используемых киберпреступником тактик, техник, найденных уязвимостей и реализуемых УБИ.

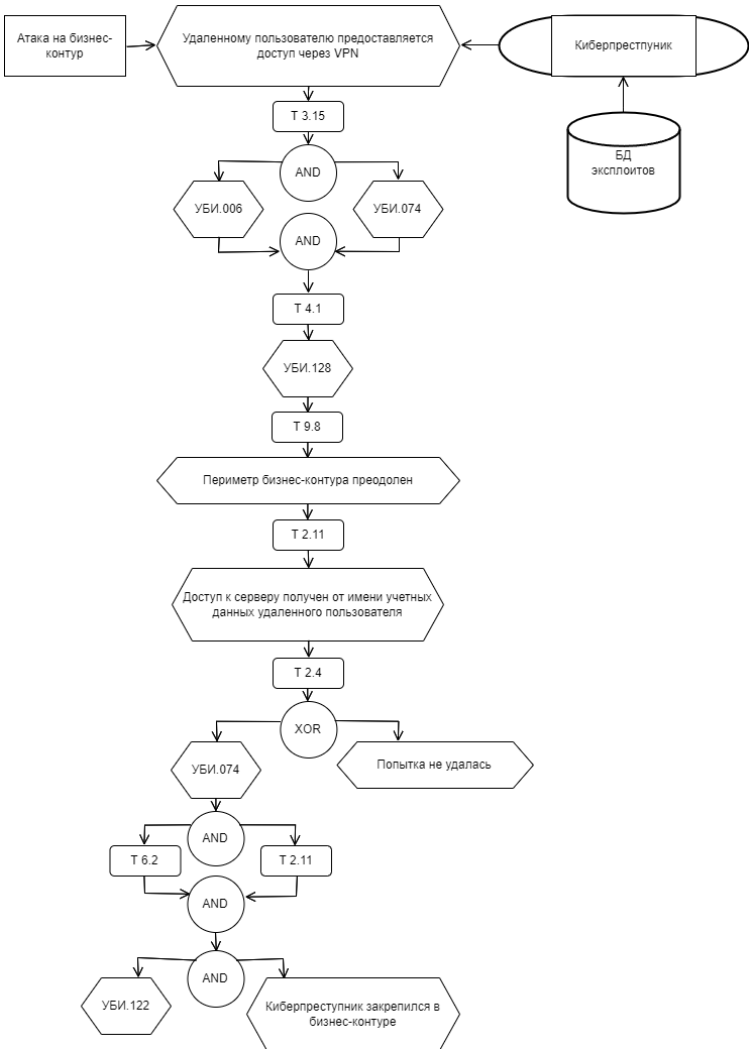



*Рис. 2. EPC-диаграмма сценария атаки на корпоративный сегмент*

Таблица 1

**Применяемые тактики, техники, реализуемые УБИ и их описание**

| **Номер тактики** | **Применяемые техники и их описание** | **Реализуемые УБИ** | **Описание** |
|---|---|---|---|
| Т3: Внедрение и исполнение вредоносного программного обеспечения в системах и сетях | Т3.15 – Планирование запуска вредоносных программ через планировщиков задач в операционной системе, а также с использованием механизмов планирования выполнения в удаленной системе через удаленный вызов процедур. Выполнение в контексте планировщика в ряде случаев позволяет авторизовать вредоносное программное обеспечение и повысить доступные ему привилегии. (CVE-2023-21541) | УБИ.006, УБИ.074 | УБИ.006: Угроза внедрения кода или данных<br>УБИ.074: Угроза несанкционированного доступа к аутентификационной информации. |

*Окончание табл. 1*

| | | | |
|---|---|---|---|
| Т4: Закрепление (сохранение доступа) в системе или сети | Т4.1 – Несанкционированное создание учетных записей или кража существующих учетных данных. (CVE-2021-38704) | УБИ.128 | УБИ.128: Угроза подмены доверенного пользователя |
| Т9: Сбор и вывод из системы или сети информации, необходимой для дальнейших действий при реализации угроз безопасности информации или реализации новых угроз | Т9.8 – Туннелирование трафика передачи данных через VPN | Событие | Периметр бизнес-контура преодолен |
| Т2: Получение первоначального доступа к компонентам систем и сетей | Т2.4 – Использование ошибок конфигурации сетевого оборудования и средств защиты, в том числе слабых паролей и паролей по умолчанию, для получения доступа к компонентам систем и сетей при удаленной атаке; Т2.11 – Несанкционированный доступ путем компрометации учетных данных сотрудника организации, в том числе через компрометацию многократно используемого в различных системах пароля (для личных или служебных нужд). (CVE-2022-23724) | УБИ.074, УБИ.122 | УБИ.074: Угроза несанкционированного доступа к аутентификационной информации. |
| | | | УБИ.122: Угроза повышения привилегий |
| Т6: Повышение привилегий по доступу к компонентам систем и сетей | Т6.2 – Подбор пароля или другой информации для аутентификации от имени привилегированной учетной записи. (CVE-2023-27121) | УБИ.122 | УБИ.122: Угроза повышения привилегий |

EPC-диаграмма сценария угрозы позволяет оценить численно вероятность реализации сценария и его ранг по следующему алгоритму:

$$P_{\text{сц}} = \prod_{i=1,j=1}^{I,J}\left[W_{CVE}(T_i) + W_{CVE}(T_j) - W_{CVE}(T_i) * W_{CVE}(T_j)\right] * \prod_{x=1,y=1}^{X,Y}\left[W_{CVE}(T_x) * W_{CVE}(T_y)\right] * 0{,}5^{K} * \prod_{l=1}^{L} W_{CVE}(T_l), \quad (1)$$

где $P_{\text{сц}}$ – вероятность реализации сценария угрозы;

$W_{CVE}(T)$ – нормированные значения уязвимостей, эксплуатируемых киберпреступником при реализации техник $T_i$, $T_j$, $T_x$, $T_y$, $T_l$;

I = J – число техник, объединенных оператором OR;

X = Y – число техник, объединенных оператором AND;

K – число техник, для которых не найдены уязвимости в базе данных (уязвимости нулевого дня);

L – число техник, каждая из которых приводит к промежуточной УБИ.

$$R_{\text{сц}} = P_{\text{сц}} * C_a,$$

где $R_{\text{сц}}$ – величина риска реализации сценария;

$C_a$ – ценность актива (объекта воздействия угрозы).

Ранг сценария определяется в зависимости от показателя риска.

Метод разработки базы знаний сценариев угроз включает в себя: правила построения EPC-диаграмм сценариев угроз, методику EPC-моделирования, оценку вероятности реализации сценария угрозы от значений эксплуатируемых уязвимостей и структуры EPC-диаграммы сценария (числа используемых логических операторов).

**Результаты численного эксперимента.** Численное значение вероятности реализации угрозы УБИ.122 рассчитывалось по формуле (1).

При этом значения эксплуатируемых при выполнении техник уязвимостей заимствованы из базы данных NIST: NVD [18–21], метрики CVSS Version 3.x. Значения уязвимостей приведены в табл. 2.

Таблица 2

**Значения уязвимостей**

| № | Наименование уязвимости | Значение Base Score | Нормированное значение |
|---|---|---|---|
| 1 | CVE-2023-21541 | 7,8 HIGH | 0,78 |
| 2 | CVE-2021-38704 | 6,1 MEDIUM | 0,61 |
| 3 | CVE-2022-23724 | 8,1 HIGH | 0,81 |
| 4 | CVE-2023-27121 | 6,1 MEDIUM | 0,61 |

В расчетах вероятность реализации техники T9.8 (туннелирование трафика передачи данных через VPN) принята равной единице, поскольку киберпреступнику удалось на предыдущих этапах многокомпонентной атаки получить права авторизованного удаленного пользователя; вероятность успешной для киберпреступника реализации техники T 2.4 принята равной 0,5, поскольку уязвимость не найдена, а результаты выполнения техники связаны оператором «исключающее ИЛИ».

Таким образом, $P_{УБИ.122}$ = 0,78* 0,61* 1* 0,81* 0,5* 0,61* 0,81 = 0,095.

Тогда численное значение риска ИБ для данной угрозы при ценности ресурса, принятой равной 0,02 составит 0,0019, то есть приближенно 0,2%.

После вычисления (для объекта воздействия угроз) вероятностей реализации всех возможных сценариев проводится их ранжирование.

**Заключение**. Предложена архитектура IRP системы, метод разработки базы знаний сценариев угроз, приведен пример построения сценария угрозы на корпоративный сегмент с промышленной сетью и результаты численного эксперимента по оценке вероятности реализации сценария. Полученные результаты позволят повысить эффективность детектирования сложных многокомпонентных атак, направленных на целевые объекты воздействия промышленной сети, будут способствовать адекватному реагированию. Функционирование модуля принятия решения по выбору варианта реагирования в составе архитектуры IRP связано с оценкой вероятности реализации сценария, его ранга; алгоритм должен обеспечивать минимизацию ущерба как от возможной атаки, так и от ответных действий.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Приказ ФСТЭК России от 14 марта 2014 № 31 «Об утверждении Требований к обеспечению защиты информации в автоматизированных системах управления производственными и технологическими процессами на критически важных объектах, потенциально опасных объектах, а также объектах, представляющих повышенную опасность для жизни и здоровья людей и для окружающей природной среды».
2. Приказ ФСТЭК России от 25 декабря 2017 № 239 «Об утверждении Требований по обеспечению безопасности значимых объектов критической информационной инфраструктуры Российской Федерации».
3. Information Security Информационная безопасность: офиц. сайт. – URL: https://www.itsec.ru/articles (дата обращения: 16.08.2024).

4. *Gaggero G.B., Armellin A., Portomauro G. and Marchese M*. Industrial Control System-Anomaly Detection Dataset (ICS-ADD) for Cyber-Physical Security Monitoring in Smart Industry Environment // IEEE Access. – 2024. – Vol. 12. – P. 64140-64149.
5. *Makrakis, Georgios Michail & Kolias, Constantinos & Kambourakis, Georgios & Rieger, Craig & Benjamin, Jacob*. Vulnerabilities and Attacks Against Industrial Control Systems and Critical Infrastructures. – 2021. – URL: https://www.researchgate.net/publication/354493711_Vulnerabilities_and_Attacks_Against_Industrial_Control_Systems_and_Critical_Infrastructures (дата обращения: 16.08.2024).
6. Jet Detective и Jet Signal внесены в реестр отечественного программного обеспечения. 12 декабря 2017. Инфосистемы Джет: офиц. сайт. – URL: https://jet.su/press-center/news/jet-detective-i-jet-signal-vneseny-v-reestr-otechestvennogo-programmnogo-obespecheniya (дата обращения: 15.08.2024).
7. Р-Вижн: офиц. сайт. – URL: https://www.rvision.ru/ (дата обращения 15.08.2024).
8. Интеллектуальная безопасность: офиц. сайт. – URL: https://www.securityvision.ru/docs/IRP.php (дата обращения: 15.08.2024).
9. Cyberbit SOC 3D. Автоматизированная консолидация всех данных управления процессом реагирования на киберинциденты на единую панель управления и повышение эффективности SOC в целом. – URL: https://www.pacifica.kz/upload/iblock/f43/f438e9f735ad1f4218e45acb9db8d706.pdf?ysclid=lzx8mckt1r891749321 (дата обращения: 16.08.2024).
10. IBM Resilient Incident Response Platform Enterprise on Cloud delivers orchestrated and automated incident response processes. IBM Asia Pacific Software Announcement AP16-0410. November 1, 2016. – URL: https://www.ibm.com/docs/en/announcements/archive/ENUSAP16-0410#abstrx (дата обращения: 15.08.2024).
11. IRP (Incident Response Platform). – URL: https://encyclopedia.kaspersky.ru/glossary/irp/ (дата обращения: 16.08.2024).
12. Методический документ ФСТЭК России от 05.02.2021. Методика оценки угроз безопасности информации. – URL: https://normativ.kontur.ru/document?moduleId=1&documentId=451500 (дата обращения: 15.08.2024).
13. *Машкина И.В., Гарипов И.Р.* Разработка EPC-моделей угроз нарушения информационной безопасности автоматизированной системы управления технологическими процессами // Безопасность информационных технологий, [S.I.]. – 2019. – Т. 26, № 4. – С. 6-20. – ISSN 2074-7136. – DOI: http://dx.doi.org/10.26583/bit.2019.4.01.
14. *Заид Алкилани М.О., Машкина И.В.* Разработка сценариев атак для оценки угроз нарушения информационной безопасности в промышленной сети // Проблемы информационной безопасности. Компьютерные системы. – 2024. – № 1 (58). – С. 96-109. – DOI 10.48612/jisp/xvkx-k619-3f2z 25.04.2024. – EDN: PDNEWN.
15. *Шеер А.В.* ARIS-моделирование бизнес-процессов. – М.: Вильямс, 2000. – 175 с.
16. Банк данных угроз безопасности информации Федеральная служба по техническому и экспортному контролю России. – URL: https://bdu.fstec.ru (дата обращения: 16.08.2024).
17. *Stouffer K., Pease M., Tang CY., Zimmerman T., Pillitteri V., Lightman S., Hahn A., Saravia S., Sherule A., Thompson M.* Title. (National Institute of Standards and Technology, Gaithersburg, MD), NIST Special Publication (SP) NIST SP 800-82r3. – 2023. – https://doi.org/10.6028/NIST.SP.800-82r3.
18. CVE-2023-21541. – https://nvd.nist.gov/vuln/detail/CVE-2023-21541 (дата обращения: 26.10.2024).
19. CVE-2021-38704. – https://nvd.nist.gov/vuln/detail/CVE-2021-38704 (дата обращения: 26.10.2024).
20. CVE-2022-23724. – https://nvd.nist.gov/vuln/detail/CVE-2022-23724 (дата обращения: 26.10.2024).
21. CVE-2023-27121. – https://nvd.nist.gov/vuln/detail/CVE-2023-27121 (дата обращения: 26.10.2024).

REFERENCES

1. Prikaz FSTEK Rossii ot 14 marta 2014 № 31 «Ob utverzhdenii Trebovaniy k obespecheniyu zashchity informatsii v avtomatizirovannykh sistemakh upravleniya proizvodstvennymi i tekhnologicheskimi protsessami na kriticheski vazhnykh ob"ektakh, potentsial'no opasnykh ob"ektakh, a takzhe ob"ektakh, predstavlyayushchikh povyshennuyu opasnost' dlya zhizni i zdorov'ya lyudey i dlya okruzhayushchey prirodnoy sredy» [Order of the FSTEC of Russia dated March 14, 2014 No. 31 "On approval of the Requirements for ensuring information security in automated control systems for production and technological processes at critical facilities, potentially hazardous facilities, as well as facilities posing an increased danger to human life and health and to the environment"].
2. Prikaz FSTEK Rossii ot 25 dekabrya 2017 № 239 «Ob utverzhdenii Trebovaniy po obespecheniyu bezopasnosti znachimykh ob"ektov kriticheskoy informatsionnoy infrastruktury Rossiyskoy Federatsii» [Order of the FSTEC of Russia dated December 25, 2017 No. 239 "On approval of the Requirements for ensuring the security of significant facilities of the critical information infrastructure of the Russian Federation"].

3. Information Security: офиц. сайт. Available at: https://www.itsec.ru/articles (accessed 16 August 2024).
4. *Gaggero G.B., Armellin A., Portomauro G. and Marchese M*. Industrial Control System-Anomaly Detection Dataset (ICS-ADD) for Cyber-Physical Security Monitoring in Smart Industry Environment, *IEEE Access,* 2024, Vol. 12, pp. 64140-64149.
5. *Makrakis, Georgios Michail & Kolias, Constantinos & Kambourakis, Georgios & Rieger, Craig & Benjamin, Jacob*. Vulnerabilities and Attacks Against Industrial Control Systems and Critical Infrastructures, 2021. Available at: https://www.researchgate.net/publication/354493711_ Vulnerabilities_and_Attacks_Against_Industrial_Control_Systems_and_Critical_Infrastructures (accessed 16 August 2024).
6. Jet Detective i Jet Signal vneseny v reestr otechestvennogo programmnogo obespecheniya. 12 dekabrya 2017. Infosistemy Dzhet: ofits. sayt [Jet Detective and Jet Signal are included in the register of domestic software. December 12, 2017. Jet Infosystems: official website]. Available at: https://jet.su/press-center/news/jet-detective-i-jet-signal-vneseny-v-reestr-otechestvennogo-programmnogo-obespecheniya (accessed 15 August 2024).
7. R-Vizhn: ofits. sayt [R-Vision: official website]. Available at: https://www.rvision.ru/ (accessed 15 August 2024).
8. Intellektual'naya bezopasnost': ofits. sayt [Intelligent security: official website]. Available at: https://www.securityvision.ru/docs/IRP.php (accessed 15 August 2024).
9. Cyberbit SOC 3D. Avtomatizirovannaya konsolidatsiya vsekh dannykh upravleniya protsessom reagirovaniya na kiberintsidenty na edinuyu panel' upravleniya i povyshenie effektivnosti SOC v tselom [Cyberbit SOC 3D. Automated consolidation of all cyber incident response management data on a single control panel and increasing the efficiency of the SOC as a whole]. Available at: https://www.pacifica.kz/upload/iblock/f43/f438e9f735ad1f4218e45acb9db8d706.pdf?ysclid=lzx8mckt1r891749321 (accessed 16 August 2024).
10. IBM Resilient Incident Response Platform Enterprise on Cloud delivers orchestrated and automated incident response processes. IBM Asia Pacific Software Announcement AP16-0410. November 1, 2016. Available at: https://www.ibm.com/docs/en/announcements/archive/ENUSAP16-0410#abstrx (accessed 15 August 2024).
11. IRP (Incident Response Platform). Available at: https://encyclopedia.kaspersky.ru/glossary/irp/ (accessed 16 August 2024).
12. Metodicheskiy dokument FSTEK Rossii ot 05.02.2021. Metodika otsenki ugroz bezopasnosti informatsii [Methodological document of the FSTEC of Russia dated 05.02.2021. Methodology for assessing information security threats]. Available at: https://normativ.kontur.ru/document?moduleId=1&documentId=451500 (accessed 16 August 2024).
13. *Mashkina I.V., Garipov I.R.* Razrabotka ERS-modeley ugroz narusheniya informatsionnoy bezopasnosti avtomatizirovannoy sistemy upravleniya tekhnologicheskimi protsessami [Development of EPC models of threats to information security of an automated process control system], *Bezopasnost' informatsionnykh tekhnologiy* [Security of Information Technology], [S.I.], 2019, Vol. 26, No. 4, pp. 6-20. ISSN 2074-7136. DOI: http://dx.doi.org/10.26583/bit.2019.4.01.
14. *Zaid Alkilani M.O., Mashkina I.V.* Razrabotka stsenariev atak dlya otsenki ugroz narusheniya informatsionnoy bezopasnosti v promyshlennoy seti [Development of attack scenarios for assessing threats of information security breach in an industrial network], *Problemy informatsionnoy bezopasnosti. Komp'yuternye sistemy* [Problems of information security. Computer systems], 2024, No. 1 (58), pp. 96-109. DOI 10.48612/jisp/xvkx-k619-3f2z 25.04.2024. – EDN: PDNEWN.
15. *Sheer A.V.* ARIS-modelirovanie biznes-protsessov [ARIS-modeling of business processes]. Moscow: Vil'yams, 2000, 175 p.
16. Bank dannykh ugroz bezopasnosti informatsii Federal'naya sluzhba po tekhnicheskomu i eksportnomu kontrolyu Rossii [Database of information security threats Federal Service for Technical and Export Control of Russia]. Available at: https://bdu.fstec.ru (accessed 16 August 2024).
17. *Stouffer K., Pease M., Tang CY., Zimmerman T., Pillitteri V., Lightman S., Hahn A., Saravia S., Sherule A., Thompson M.* Title. (National Institute of Standards and Technology, Gaithersburg, MD), NIST Special Publication (SP) NIST SP 800-82r3, 2023. Available at: https://doi.org/10.6028/NIST.SP.800-82r3.
18. CVE-2023-21541. Available at: https://nvd.nist.gov/vuln/detail/CVE-2023-21541 (accessed 26 October 2024).
19. CVE-2021-38704. Available at: https://nvd.nist.gov/vuln/detail/CVE-2021-38704 (accessed 26 October 2024).

20. CVE-2022-23724. Available at: https://nvd.nist.gov/vuln/detail/CVE-2022-23724 (accessed 26 October 2024).
21. CVE-2023-27121. Available at: https://nvd.nist.gov/vuln/detail/CVE-2023-27121 (accessed 26 October 2024).

Статью рекомендовал к опубликованию д.т.н., профессор А.Н. Целых.

**Машкина Ирина Владимировна –** Уфимский университет науки и технологий; e-mail: profmashkina@mail.ru; г. Уфа, Россия; тел.: +79279277089; кафедра вычислительной техники и защиты информации; д.т.н.; профессор.

**Уразаева Айгуль Маратовна** – e-mail: ajgul.urazaeva2017@yandex.ru; кафедра вычислительной техники и защиты информации; студент.

**Mashkina Irina Vladimirovna** – Ufa University of Science and Technology; e-mail: profmashkina@mail.ru; Ufa, Russia; phone: +79279277089; the Department of Computer Science and Information Security; dr. of eng. sc.; professor.

**Urazaeva Aigul Maratovna** – e-mail: ajgul.urazaeva2017@yandex.ru; the Department of Computer Science and Information Security; student.

УДК 004.853 DOI 10.18522/2311-3103-2024-5-88-102

**Е.М. Герасименко, Ю.А. Кравченко, Д.А. Шаненко**

# АЛГОРИТМ ПОИСКА И ПРИОБРЕТЕНИЯ ЗНАНИЙ НА ОСНОВЕ ТЕХНОЛОГИЙ ОБРАБОТКИ И АНАЛИЗА ТЕКСТОВ НА ЕСТЕСТВЕННОМ ЯЗЫКЕ*

*Статья посвящена решению актуальной научной проблемы повышения эффективности обработки и анализа текстовой информации при решении задач поиска и приобретения знаний. Актуальность данной задачи связана с необходимостью создания эффективных средств обработки накапливаемого огромного количества слабо структурированных данных, содержащих важные, иногда скрытые знания, необходимые для построения эффективных систем управления сложными объектами различной природы. Предлагаемый автором алгоритм поиска и приобретения знаний при обработке и анализе текстовой информации, отличается применением низкоуровневых детерминированных правил, позволяющих провести качественное упрощение текста на основе исключения из текстовой информации слов, инвариантных к смыслу. Алгоритм опирается на доменную проработку, позволяющую сформировать списки доменно-специфичных слов, что позволяет обеспечить высокое качество упрощения текста. В данной задаче исходными данными являются потоки текстовой информации (описание профилей), извлеченных из онлайн платформ для рекрутинга, выходная информация представляется предложениями, сформированными в виде тройки «субъект-глагол-объект», отражающих гранулы знаний, полученных в процессе обработки текста. Использование данного порядка единиц, составляющих предложение, обусловлено тем фактом, что данный порядок наиболее распространен в русском языке, хотя в самих текстах возможны иные вариации порядка без потери общего смысла. Основная идея алгоритма заключается в разбиении большого корпуса текста на предложения с последующей фильтрацией полученных предложений на основании введенных пользователем ключевых слов. В последствии предложения разделяются на компоненты и упрощаются в зависимости от вида поступившей компоненты (глагольная, именная). В качестве примера в данной работе использовалась сфера маркетинга, а ключевыми словами выступили «социальные сети». Автором разработан алгоритм поиска и приобретения знаний на основе технологий обработки и анализа текстов на естественном языке, а также была выполнена программная реализация предложенного алгоритма. В качестве методов оценки эффективности использовался ряд метрик: индекс Флэша-Кинкейда; индекс Кол-*

* Исследование выполнено за счет гранта Российского научного фонда № 22-71-10121, https://rscf.ru/project/22-71-10121/ в Южном федеральном университете.

*ман-Лиау; автоматический индекс удобочитаемости. Проведенные вычислительные эксперименты подтвердили эффективность предложенного алгоритма по сравнению с аналогами, использующими нейронные сети для решение подобных задач.*

*Обработка текстовой информации; поиск знаний; извлечение знаний; обработка естественного языка; извлечение компонент; упрощение текста.*

**E.M. Gerasimenko, Yu.A. Kravchenko, D.A. Shanenko**

## ALGORITHM FOR SEARCHING AND ACQUISITION OF KNOWLEDGE BASED ON TECHNOLOGIES FOR PROCESSING AND ANALYZING TEXTS IN NATURAL LANGUAGE

*The article is devoted the topical scientific problem of increasing the efficiency of processing and analyzing text information when solving problems of searching and acquiring knowledge. The relevance of this task is related to the need to create effective means of processing the accumulated huge amount of poorly structured data containing important, sometimes hidden knowledge that is necessary for building effective control systems for complex objects of different nature. The algorithm of search and knowledge acquisition in processing and analyzing textual information proposed by the author is characterized by the use of low-level deterministic rules that allow for qualitative text simplification based on the exclusion of words invariant to meaning from textual information. The algorithm relies on domain elaboration that allows to create lists of domain-specific words, which allows for high quality text simplification. In this task, the input data are streams of textual information (profile descriptions) extracted from online recruiting platforms; the output information is represented by sentences formed in the form of a triple "subject-verb-object", reflecting the granules of knowledge obtained during text processing. The use of this order of units constituting a sentence is due to the fact that this order is the most widespread in the Russian language, although other variations of the order are possible in the texts themselves without losing the general meaning. The main idea of the algorithm is to split a large corpus of text into sentences, then filter the resulting sentences based on the keywords entered by the user. Subsequently, the sentences are further split into components and simplified depending on the type of received component (verbal, nominal). The field of marketing was used as an example in this work, and the keywords were "social media". The author has developed an algorithm for for knowledge search and acquisition based on natural language text processing and analysis technologies, and a software implementation of the proposed algorithm has been performed. A number of metrics were used as efficiency evaluation methods: the Flash-Kincaid index; the Coleman-Liau index; and the automatic readability index. The conducted computational experiments have confirmed the effectiveness of the proposed algorithm in comparison with analogues that use neural networks to solve similar problems.*

*Text information processing; knowledge retrieval; knowledge extraction; natural language processing; component extraction; text simplification.*

**Введение.** Задачи поиска знаний при обработке и анализе текстовой информации востребованы во многих типах прикладных информационных систем. Традиционные методы поиска знаний основаны на правилах, тезаурусах, машинном обучении с учителем и требуют наличия достаточных априорных знаний об исследуемой предметной области в виде лингвистических ресурсов, обработанных корпусов текста, словарей и грамматик.

Для создания правил и грамматик отдельных предметных областей разработано достаточное количество инструментов, которые постоянно применяются в информационных системах: CPSL [1]; Jape [2]; UIMA Ruta [3]; ABBYY Compreno [4] и др.

Основным недостатком перечисленных инструментов являются значительные трудозатраты на разработку правил и грамматик.

Поиск и извлечение знаний играют решающую роль в области искусственного интеллекта, позволяя компьютерам понимать данные на уровне, сравнимом с человеком [5]. Также они помогают извлекать ценные сведения, обнаруживать скрытые закономерности и придавать смысл большим объемам данных. Все это может быть использовано в различных отраслях, таких как здравоохранение, финансы, производство и др., для стимулирования инноваций, совершенствования процессов и получения конкурентоспособных преимуществ.

Извлечение знаний включает в себя ряд этапов, направленных на извлечение и преобразование данных в значимые единицы знаний. Эти этапы обычно включают предварительную обработку данных, выбор функций, интеллектуальный анализ данных (data mining) и представление знаний.

Спектр методов, используемых для решения задач извлечения знаний, достаточно обширен [6–14]. Это относительно новая тенденция, возникшая в связи с необходимостью масштабирования систем извлечения текстовой информации до очень больших объемов данных, которые представляют собой смесь огромного количества различных тем.

Актуальность работы связана с необходимостью создания эффективных средств обработки накапливаемого огромного количества слабо структурированных данных, содержащих важные, иногда скрытые знания, необходимые для построения эффективных систем управления сложными объектами различной природы.

**1. Постановка задачи.** При обработке естественного языка важно выявить и извлечь наиболее релевантные и значимые слова (или фразы) из заданного текста. В данном процессе главную роль играют парсеры, которые анализируют синтаксическую структуру текста и помогают выявить ключевые компоненты, отражающие важные понятия.

*Синтаксический анализ структуры* позволяет сосредоточиться на конкретных синтаксических единицах при извлечении знаний из текстов. Подобные единицы часто несут в себе важную информацию о субъекте, объекте или действии [15].

*Синтаксический анализ зависимостей* устанавливает отношения между подлежащим, сказуемым и дополнением, а также находит другие значимые синтаксические зависимости, которые вносят вклад в общий смысл предложения (например, в сложноподчиненных предложениях придаточные дополняют главное) [16–21]. Из данных зависимостей можно извлечь знания, особенно если слова играют важную роль в структуре предложения.

*Упрощение предложений.* Синтаксический анализ позволяет получить сведения о конкретных синтаксических единицах, но зачастую они могут быть нагружены связанными с ними «дополнениями» (например, прилагательными или наречиями) [22–27]. Попытка работать с полноценным предложением, извлечь из него элементы знаний, занимает больше времени, по сравнению с упрощенными компонентами.

Поэтому для повышения эффективности обработки текстовой информации требуется представить следующие решения:

1) определить наиболее подходящую структуру алгоритма и реализацию модулей алгоритма;

2) реализовать алгоритм поиска и приобретения знаний, наиболее подходящий под условия дальнейшей эксплуатации.

Оценка эффективности работы алгоритма будет проводиться при помощи следующих метрик:

- Индекс Флэша-Кинкейда.
- Индекс Колман-Лиау.
- Автоматический индекс удобочитаемости.

Опишем основные метрики более подробно.

**Индекс Флэша-Кинкейда.** Применяется для измерения уровня сложности текста на основе длины слов и предложений.

Значения варьируются от 0 до 18, где 18 соответствует наиболее сложному тексту. Расчет индекса производится по формуле:

$$F_1 = FKGLF = 0.39\frac{total\ words}{total\ sentences} + 11.8\frac{total\ syllables}{total\ words} - 15.59, \quad (1)$$

где total words – общее количество слов в тексте, total sentences – общее количество предложений в тексте, a total syllables – общее количество символов текста.

Индекс по шкале принято распределять следующим образом:

а) 0 – 1 – текст очень легко читается;

б) 1 – 5 – текст легко читается;

в) 5 – 11 – достаточно легко читается. Стандартный разговорный язык;

г) 11 – 18 – текст сложен для восприятия;

д) >18 – текст очень трудно читается.

**Индекс Колман-Лиау.** Основывается на анализе количества букв в слове и слов в предложении, опирается на количество символов, а не слогов в слове, поскольку символы легче и точнее подсчитываются компьютерными программами, чем слоги. Расчет индекса производится по формуле:

$$F_2 = CLI = 0.0588\mathrm{L} - 0.296S - 15.8, \quad (2)$$

где L – среднее количество букв на 100 слов, а S – среднее количество предложений на 100 слов.

Индекс по шкале принято распределять следующим образом:

а) 0 – 1 – текст очень легко читается;

б) 1 – 5 – текст легко читается;

в) 5 – 8 – достаточно легко читается. Стандартный разговорный язык;

г) 8 – 11 – текст сложен для восприятия;

д) >11 – текст очень трудно читается.

**Автоматический индекс удобочитаемости.** В отличие от индекса Колман-Лиау, подсчитывает не слоги, а символы. От количества символов зависит сложность слова, помимо этого ведется подсчет предложений. Расчет индекса производится по формуле:

$$F_3 = ARI = 4.71 \times \frac{C}{W} + 0.5 \times \frac{W}{S} - 21.43, \quad (3)$$

где C – среднее количество букв и цифр в тексте, W – количество слов в тексте, S – количество предложений в тексте.

Индекс по шкале принято распределять следующим образом:

а) 0 – 1 – текст очень легко читается;

б) 1 – 5 – текст легко читается;

в) 5 – 8 – достаточно легко читается. Стандартный разговорный язык;

г) 8 – 11 – текст сложен для восприятия;

д) >11 – текст очень трудно читается.

Целевой функцией алгоритма является минимизация значений описанных мер читабельности текста.

$$F_i \rightarrow min\,,$$

где i – номер метрики.

Каждая из представленных мер читабельности имеет порог, после которого текст становится очень трудным для восприятия (нечитабельным). Показатели удобочитаемости являются алгоритмическими эвристиками, используемыми для оценки удобочитаемости. Стоит отметить, что удобочитаемость не является показателем качества письма и что эти эвристические методы являются лишь оценкой понятности отрывка.

**2. Обобщенная схема процесса поиска и приобретения знаний.** Для проведения вычислительных экспериментов была выбрана предметная область, связанная с установлением деловых контактов. Предложенная авторами версия алгоритма представляет собой набор модулей для использования в информационно-поисковой системе в сфере маркетинговых исследований.

Обобщенная схема процесса обработки текста алгоритмами представлена на рис. 1.

На обработку поступает корпус текста. В самом начале текст разбивается на отдельные предложения, после этого происходит фильтрация получившихся предложений по заданным ключевым словам. После фильтрации на дальнейшую обработку поступают только отобранные на предыдущем шаге предложения. Если предложения сложные (имеют в себе более одной пары подлежащего и сказуемого), то они дополнительно разделяются на простые предложения с сохранением отношений (сохраняется информация о том, что предложение было сложносочиненными или сложноподчиненными). После этого предложения разделяются на ключевые и дополняющие компоненты. На заключительном этапе происходит упрощение компонент (вырезаются прилагательные, наречия). После этого полученные результаты работы алгоритма вносятся в базу данных и доступны для дальнейшей работы.


*Рис. 1. Обобщение процесса обработки текстов*

На рис. 2 представлена схема модуля разбиения текста на предложения.

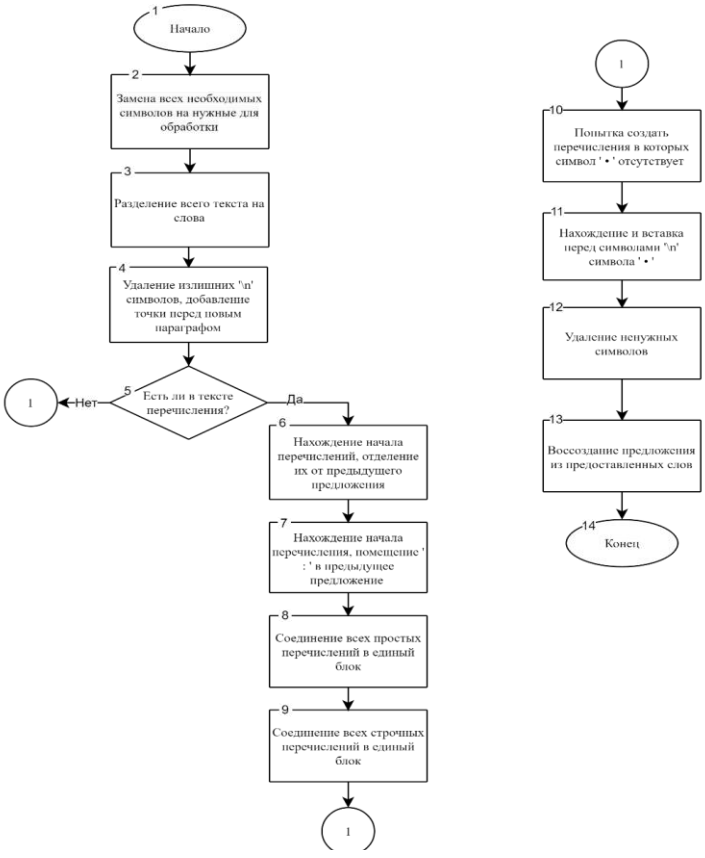



*Рис. 2. Схема разбиения текста на предложения*

Модуль фильтрации предложений принимает в качестве входных данных список текстов (предложений), список ключевых слов и параметр to_lowercase (необязателен, установлено значение «Истина» по умолчанию для перевода слов в нижний регистр). Он перебирает список текстов и перечисляет их (т.е. присваивает каждому тексту идентификаторы). Для каждого текста вызывает функцию поиска ключевых слов, чтобы попытаться получить список ключевых слов. Если ключевые слова найдены, создает экземпляр специального файла, передает ему соответствующую информацию (текст, найденные ключевые слова и идентификатор текста) и возвращает список экземпляров специального файла.

На рис. 3 представлена схема модуля фильтрации предложений. Модуль разбиения сложных предложений опирается на полученные от spaCy отношения зависимостей. Отношения зависимостей, используемые для идентификации и разделения предложений: cconj для сложносочинённых и acl:relcl, acl, xcomp, ccomp для сложноподчиненных предложений.


*Рис. 3. Схема фильтрации предложений*

Сначала определяется корень предложения (подразумевается, что этим словом является глагол), а затем проверяется наличие зависимости между корнем и другими частями предложения. В случае если корнем предложения стала другая часть речи, проводится обработка неправильного корня слова, в процессе которой проверяется наличие у корня слова зависимостей вида acl, relcl или conj. Функция работает рекурсивно: сначала предложения разделяются на одном уровне (разбиваются по ширине), а затем происходит переход на более глубокий уровень (разбиваются по глубине). Рекурсия вызвана тем фактом, что сложные предложения также могут быть составными (сложносочиненное предложение будет главным для зависимого).

На рис. 4 представлена схема модуля разбиения сложных предложений.

Модуль деления на компоненты разделяет части простого предложения, полученного на предыдущем этапе, на определенные грамматические компоненты на основе зависимостей spaCy. Поскольку предложение представлено в виде глагола и его зависимых элементов, задача рассматривается как присвоение определенных меток зависимым от глагола элементам.

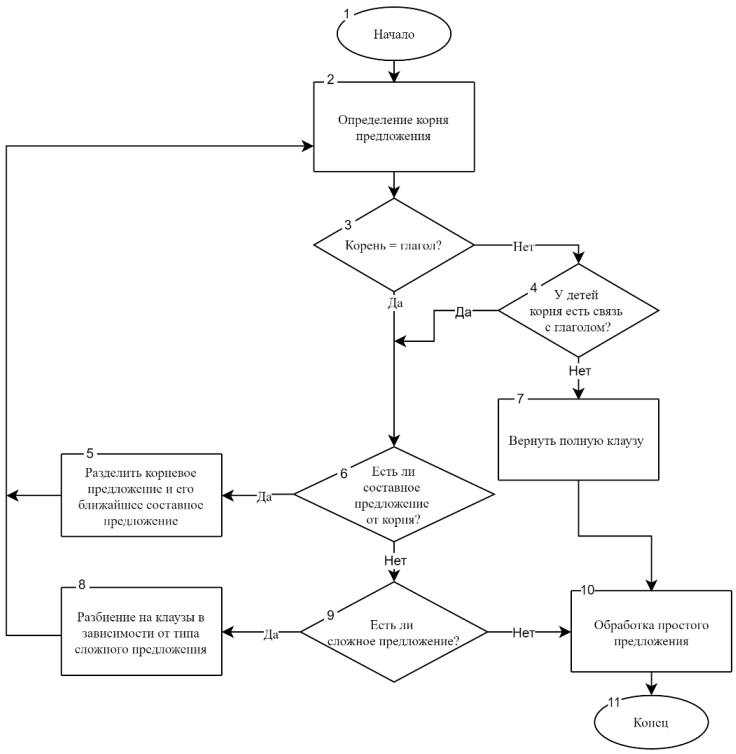



*Рис. 4. Схема разбиения сложных предложений*

На рис. 5 представлена схема модуля деления на компоненты предложений.

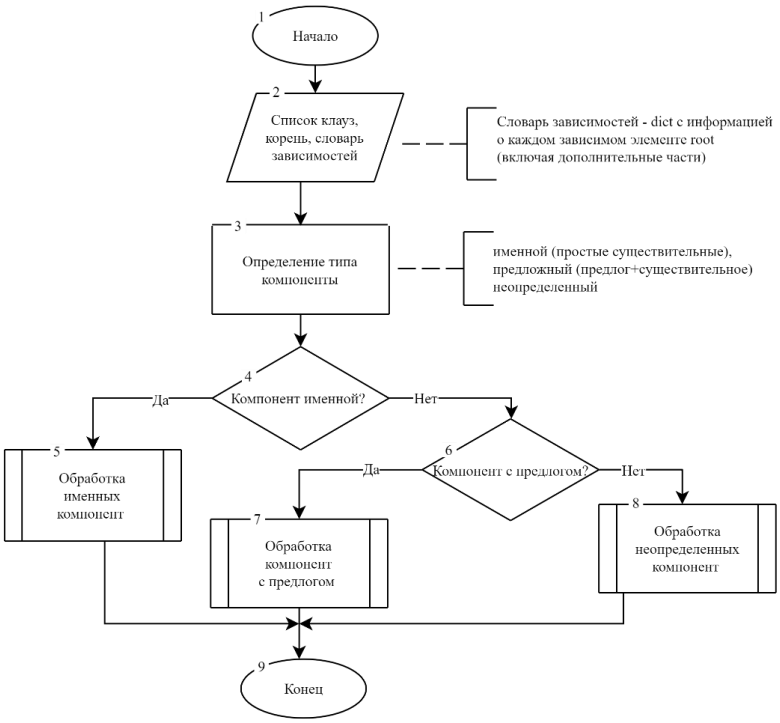



*Рис. 5. Схема деления на компоненты*

Модуль упрощения предназначен для упрощения глагольных и именных компонент текста. Он оценивает каждый компонент на основе его длины, типа и специфических лингвистических особенностей, а затем классифицирует составляющие его лексемы на основные (ядро) и дополнительные (детали) элементы.

На рис. 6 представлена схема общего процесса упрощения компонент предложения.

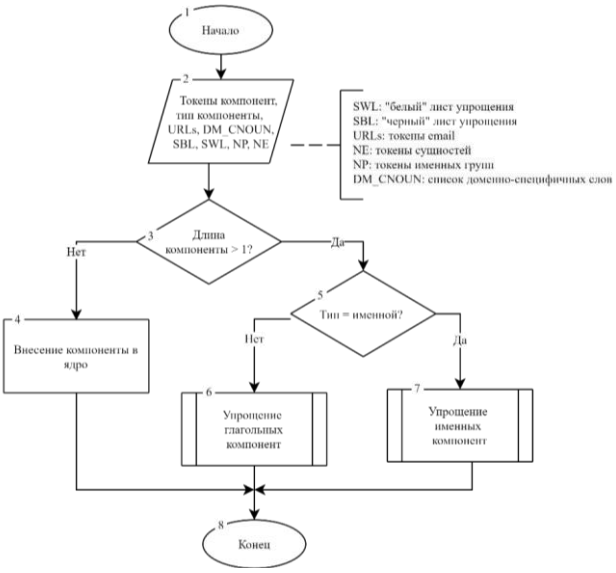



*Рис. 6. Схема общего процесса упрощения компонент*

Сначала определяется длина компоненты. Если длина больше 1, то упрощение не выполняется. Если компонента имеет тег части речи «глагол», производится переход к упрощению глагольного компонента; в противном случае переход к упрощению именного компонента.

На рис. 7 представлена схема упрощения именных компонент, а на рис. 8 – глагольных компонент.

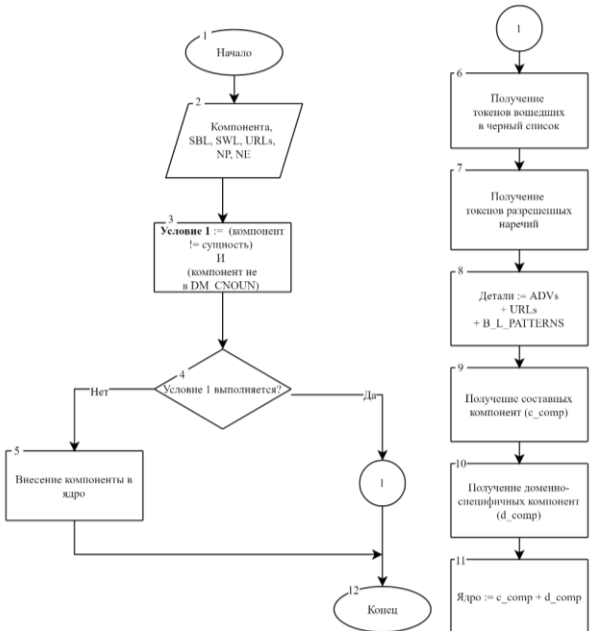



*Рис. 7. Схема упрощения именных компонент*

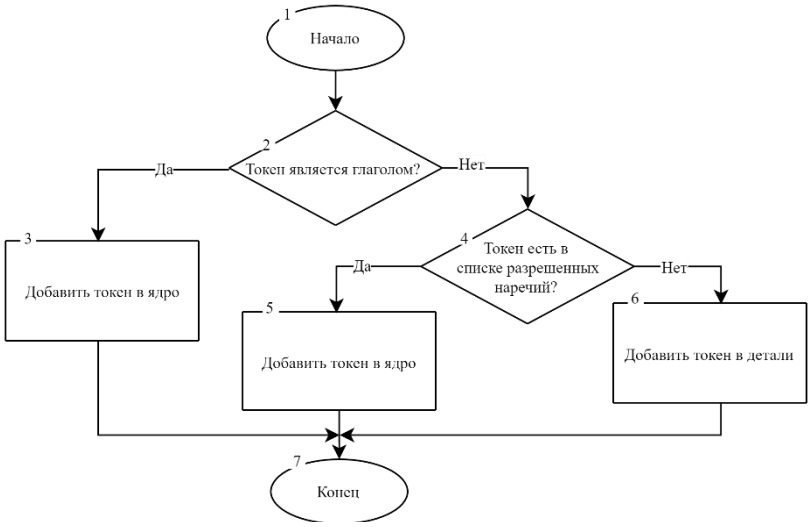



*Рис. 8. Схема упрощения глагольных компонент*

При упрощении именного компонента выполняется ряд проверок: если компонент не является сущностью, не внесен в «белый» список для упрощения, то выполняется упрощение именных компонент при помощи фильтрации неважных для передачи смысла предложения токенов на основе заранее определенных критериев.

Для каждого токена в глагольной компоненте выполняется следующая проверка: если токен является глаголом, то происходит добавление его в ядро. В противном случае – добавление токена в раздел «Детали».

**3. Компонентная архитектура разработанного программного приложения и вычислительный эксперимент.** На рис. 9 показана архитектура модулей разработанного программного приложения.

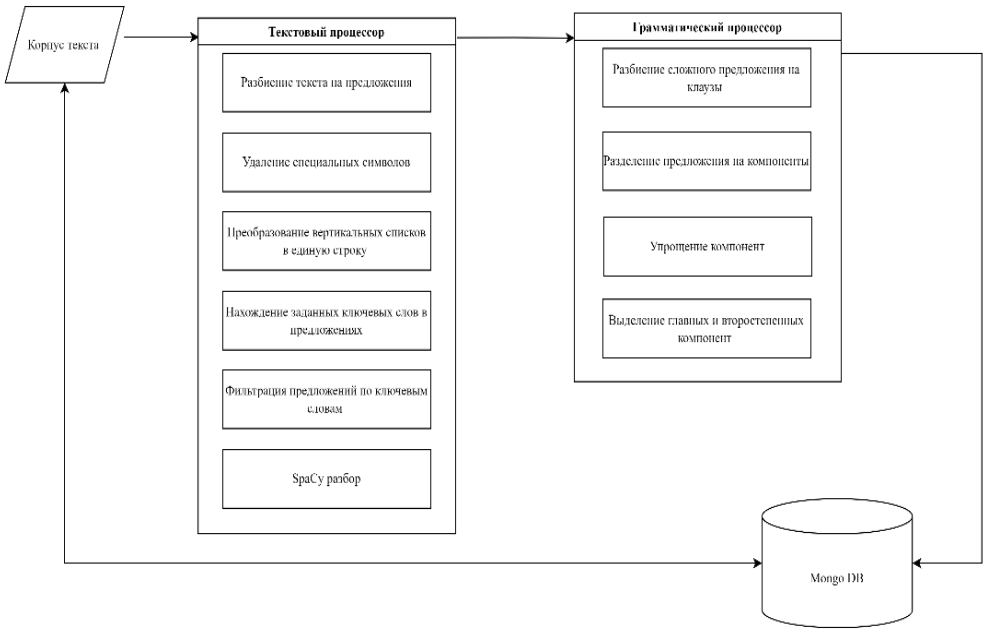



*Рис. 9. Схема архитектуры модулей*

Алгоритм направлен на упрощение предложений для увеличения скорости передачи и обработки информации: поступающий на вход корпус текста будет проходить через предварительный этап обработки, где поступивший текст будет разбит на отдельные предложения, из которых удаляются специальные символы. После этого будут отбираться те предложения, которые подходят под заданные ключевые слова.

Предложения последовательно будут отправляться на обработку spaCy. Далее предложение будет разбито на клаузы (клауза – группа слов, состоящая из подлежащего и глагола; элементарное предложение), а после этого каждая клауза будет разделена на компоненты. После этого будет запущен модуль упрощения. Здесь будут выделяться важные элементы внутри компонент, поскольку именно они будут представлять извлеченные знания. Все прилагательные и наречия, не попавшие в списки допуска, будут помечаться как детали и подразумеваться как слова, не имеющие критической важности для смысла.

После этого предложения с упрощенными элементами будут сохранены в БД и доступны для дальнейшей работы.

Для исследования работы алгоритма будем проверять качество упрощения предложений с использованием сразу нескольких метрик:

- индекса Флэша-Кинкейда, в котором значения варьируются от 0 до 18, где 18 соответствует наиболее сложному тексту;
- индекса Колман-Лиау, в котором значения варьируются от 0 до 11, где 11 соответствует наиболее сложному тексту;
- автоматического индекса удобочитаемости, в котором значения варьируются от 0 до 11, где 11 соответствует наиболее сложному тексту.

Проверим сложность чтения до работы алгоритма и после, а также сравним с результатами аналогов. Поскольку предложения являются сложными, то значения метрики будут превышены для всех рассматриваемых экспериментов. Цель – максимально приблизиться к верхним пороговым значениям.

Тестироваться будет следующее предложение:

«Social Island является рекламным агентством, которое использует силу психологического воздействия для своих любимых клиентов, чтобы обеспечить мгновенную вирусность в социальных сетях».

Авторское решение разделяет предложения на клаузы и работает с ними, поэтому результат будет отражен в виде последовательности полных клауз, полученных в результате работы. Результаты упрощения показаны в табл. 1.

Таблица 1

**Результаты упрощения предложения**

| Сервис «Robotext» | Сервис «Neural Writer» | Авторское решение |
|---|---|---|
| Social Island – рекламное агентство, которое использует силу психологического воздействия для своих любимых клиентов, чтобы обеспечить мгновенную вирусность в социальных сетях. | Social Island использует психологическое воздействие, чтобы реклама стала вирусной в социальных сетях. | Social Island является рекламным агентством; агентство использует силу воздействия для клиентов; агентство обеспечить вирусность в социальных сетях. |

Сервис Neural Writer полностью изменил предложение, не сохранив его структуру, а вот Robotext практически не изменил предложение.

Учитывая, что объем информации растет экспоненциально, не всегда системы позволяют получить ответ за допустимое время. Проверим насколько изменилась читабельность тестового предложения, а также сравним время отклика при помощи формул (1), (2), (3). Результаты сведены в табл. 2.

Таблица 2

**Оценка сложности предложений и времени обработки**

| Метрика | Изначальное предложение | Сервис «Robotext» | Сервис «Neural Writer» | Авторское решение |
|---|---|---|---|---|
| Индекс Флэша-Кинкейда | 21.77 | 21.68 | 21.27 | 17.57 |
| Индекс Колман-Лиау | 23.25 | 22.91 | 22.84 | 18.73 |
| Автоматический индекс удобочитаемости | 25.83 | 25.28 | 25.04 | 22.58 |
| Скорость работы, сек. | – | 1.8 | 2.4 | 1.9 |

Согласно индексу Флэша-Кинкейда, авторский алгоритм сумел упростить предложение до индекса 16, что соответствует тексту, который сложно читается в отличие от начального индекса, который показал, что предложение очень трудное для восприятия. Индекс Колмана-Лиац, а также автоматический индекс читаемости все еще превышают верхние пороговые значения показателей.

Проверим сложность 100 текстовых предложений, выведем средние показатели сложности в табл. 3.

Таблица 3

**Средняя оценка сложности 100 предложений**

| Метрика | Изначальное предложение | Сервис «Robotext» | Сервис «Neural Writer» | Авторское решение |
|---|---|---|---|---|
| Средний индекс Флэша-Кинкейда | 21.77 | 21.69 | 19.54 | 18.38 |
| Средний индекс Колман-Лиау | 23.25 | 20.64 | 19.42 | 16.31 |
| Средний автоматический индекс удобочитаемости | 25.83 | 23.8 | 21.72 | 18.32 |
| Средняя скорость работы на одном предложении, сек | – | 2.08 | 2.35 | 2.12 |

Табл. 3 демонстрирует, что упор на проработку домена позволяет добиться лучших показателей упрощения текста при допустимых временных рамках.

Временная сложность определяет, сколько времени требуется алгоритму для решения задачи при увеличении объема входных данных. Обычно она оценивается по числу операций, которые необходимо выполнить алгоритму.

Нотация Big-O позволяет показать эффективность (или быстродействие) алгоритма. Big-O позволяет легко сравнивать скорости работы алгоритмов и дает общее представление о том, сколько времени потребуется алгоритму для выполнения всех требуемых от него действий [28]. Термин «Big-O» обычно используется для описания общей производительности, но на самом деле описывает «наихудшую» (т.е. самую низкую) скорость, с которой может работать алгоритм.

Поскольку алгоритм использует различные операции (вложенные циклы, операция sort(), рекурсия), согласно [29], временная сложность алгоритма составляет в худшем случае: $O(n^2)$, т.е. представляет квадратичную сложность.

**Заключение.** В данной статье описана разработка алгоритма поиска и приобретения знаний. Основной целью работы являлось повышении эффективности алгоритмов поиска и приобретения знаний на основе технологий обработки и анализа текстов на естественном языке.

Основными результатами проведенного исследования являются следующие:

1. Представлена формализованная постановка решаемой задачи.

2. Разработан алгоритм поиска и приобретения знаний, которая отличается от известных аналогов использованием доменной проработки и применением детерминированных низкоуровневых правил, позволяющих обеспечить высокий уровень упрощения анализируемой текстовой информации.

Для оценки эффективности предложенного алгоритма разработано программное приложение и проведен вычислительный эксперимент. Полученные результаты проведенных экспериментальных исследований подтверждают эффективность предложенного алгоритма поиска и приобретения знаний на основе технологий обработки и анализа текстов на естественном языке. Временная сложность представленного алгоритма является квадратичной.

Данный алгоритм повышает эффективность обработки текстовой информации в различных доменах, требующих обработки больших объёмов данных, позволяя уменьшить сложность текста.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК


1. *Appelt D.E.* The common pattern specification language // Technical report / SRI International, Artificial Intelligence Center. – 1998.
2. *Cunningham H., Maynard D., Bontcheva K., Tablan V.* A framework and graphical development environment for robust NLP tools and applications // Proceedings of the 40th Anniversary Meeting of the Association for Computational Linguistics. – 2002. – P. 168-175.
3. *Kluegl P., Toepfer M., Beck P.D. et al.* UIMA Ruta: Rapid development of rule-based information extraction applications // Natural Language Engineering. – 2016. – Issue 22, No. 1. – P. 1-40.
4. *Starostin A.S., Smurov I.M., Stepanova M.E.* A production system for information extraction based on complete syntactic semantic analysis // Papers from the Annual International Conference "Dialogue". – 2014. – P. 659-667.
5. *Куршев Е.П., Сулейманова Е.А., Трофимов И.В.* Роль знаний в системах извлечения информации из текстов // Программные системы: теория и приложения. – 2012. – Т. 3, № 3. – С. 57-70.
6. *Blanko M., Cafarella M.J., Soderland S. et al.* Open information extraction from the web // Proceedings of the 20th International Joint Conference on Artifical Intelligence. – 2007. – P. 2670-2676.
7. *Banko M., Etzioni O.* The tradeoffs between open and traditional relation extraction // Proceedings of ACL-08: HLT. – 2008. – P. 28-36.
8. *Zhu J., Nie Z., Liu X. et al.* StatSnowball: a statistical approach to extracting entity relationships // Proceedings of the 18th international conference on World wide web. – 2009. – P. 101-110.
9. *Wu F., Weld D.S.* Open information extraction using Wikipedia // Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics. – 2010. – P. 118-127.
10. *Fader A., Soderland S., Etzioni O.* Identifying relations for open information extraction // Proceedings of the Conference on Empirical Methods in Natural Language Processing. – 2011. – P. 1535-1545.
11. *Etzioni O., Fader A., Christensen J. et al.* Open information extraction: The second // Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence. – 2011. – P. 3-10.
12. *Schmitz M., Bart R., Soderland S. et al.* Open language learning for information extraction // Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning. – 2012. – P. 523-534.
13. *Angeli G., Johnson Premkumar M.J., Manning C.D.* Leveraging linguistic structure for open domain information extraction // Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing. – 2015. – P. 344-354.
14. *Nakashole N., Weikum G., Suchanek F.* PATTY: A taxonomy of relational patterns with semantic types // Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning. – 2012. – P. 1135-1145.
15. *Амурская О.Ю., Егорова А.Д.* Синтаксический парсинг как анализ структуры предложения (синтагмы) // Филология и культура. – 2022. – № 4 (70). – С. 14-23.

16. *Кобзарева Т.Ю.* Лингвистический базис анализа поверхностно-синтаксических связей сегментов в русском предложении // Вестник РГГУ. Серия: История. Филология. Культурология. Востоковедение. – 2008. – № 6. – С. 157-170.
17. *Кобзарева Т.Ю.* Иерархия задач поверхностно-синтаксического анализа русского предложения // Научно-техническая информация. Серия 2: Информационные процессы и системы. – 2007. – № 1. – С. 23-35.
18. *Кобзарева Т.Ю., Лахути Д.Г., Ножов И.М.* Сегментация русского предложения: поверхностно-синтаксический анализ как самостоятельный модуль анализа текста // Матер 5-ой международной конференции «Информационное общество, информационные ресурсы и технологии телекоммуникации». Секция «Интеллектуальные системы автоматизированной поддержки научных исследований». – М.: ВИНИТИ. НТИ, 2000. – С. 31-34.
19. *Кобзарева Т.Ю.* Проблема кореференции в рамках поверхностно-синтаксического анализа русского языка // Компьютерная лингвистика и интеллектуальные технологии: Тр. Международной конференции «Диалог 2003». – М.: Наука, 2003. – С. 278 -284.
20. *Кобзарева Т.Ю.* Принципы сегментационного анализа русского предложения // Московский лингвистический журнал. – 2004. – Т. 8, № 1. – С. 31-80.
21. *Кобзарева Т.Ю.* Поиск хозяина предложной группы в русском предложении // Компьютерная лингвистика и интеллектуальные технологии: Тр. Международной конференции «Диалог-2010». – 2010. – Вып. 9 (16). – С. 186-191.
22. *Rets I., Astruc L., Coughlan T., Stickler U.* Approaches to simplifying academic texts in English: English teachers' views and practices // English for Specific Purposes. – 2022. – Issue 68. – P. 31-46.
23. *Crossley S.A.*, *Allen D., McNamara D.S.* Text simplification and comprehensible input: A case for an intuitive approach // Language Teaching Research. – 2012. – Issue 16, No. 1. – P. 89-108. – DOI: 10.1177/1362168811423456.
24. *Soemer A., Schiefele U.* Text difficulty, topic interest, and mind wandering during reading // Learning and Instruction. – 2019. – Issue 61, No. 1. – P. 12-22.
25. *Allen D.* A study of the role of relative clauses in the simplification of news texts for learners of English // System. – 2009. – Issue 37, No. 4. – P. 585-599.
26. *Long M.H.* Optimal input for language learning: Genuine, simplified, elaborated, or modified elaborated? // Language Teaching. – 2020. – Issue 53, No. 2. – P. 169-182. – DOI: 10.1017/S0261444819000466.
27. *Tickoo M.L.* Simplification: Theory and Application. Anthology Series 31 // ERIC Clearinghouse. – 1993. – P. 254.
28. Big-O // Big-O.io. – URL: https://big-o.io (дата обращения: 03.07.2024).
29. Complexity Cheat Sheet for Python Operations // GeeksforGeeks. – URL: https://www.geeksforgeeks.org/complexity-cheat-sheet-for-python-operations/ (дата обращения: 03.07.2024).

REFERENCES

1. *Appelt D.E.* The common pattern specification language, *Technical report,* SRI International, Artificial Intelligence Center, 1998.
2. *Cunningham H., Maynard D., Bontcheva K., Tablan V.* A framework and graphical development environment for robust NLP tools and applications, *Proceedings of the 40th Anniversary Meeting of the Association for Computational Linguistics,* 2002, pp. 168-175.
3. *Kluegl P., Toepfer M., Beck P.D. et al.* UIMA Ruta: Rapid development of rule-based information extraction applications, *Natural Language Engineering,* 2016, Issue 22, No. 1, pp. 1-40.
4. *Starostin A.S., Smurov I.M., Stepanova M.E.* A production system for information extraction based on complete syntactic semantic analysis, *Papers from the Annual International Conference "Dialogue",* 2014, pp. 659-667.
5. *Kurshev E.P., Suleymanova E.A., Trofimov I.V.* Rol' znaniy v sistemakh izvlecheniya informatsii iz tekstov [The role of knowledge in systems of information extraction from texts], *Programmnye sistemy: teoriya i prilozheniya* [Software systems: theory and applications], 2012, Vol. 3, No. 3, pp. 57-70.
6. *Blanko M., Cafarella M.J., Soderland S. et al.* Open information extraction from the web, *Proceedings of the 20th International Joint Conference on Artifical Intelligence,* 2007, pp. 2670-2676.
7. *Banko M., Etzioni O.* The tradeoffs between open and traditional relation extraction, *Proceedings of ACL-08: HLT,* 2008, pp. 28-36.
8. *Zhu J., Nie Z., Liu X. et al.* StatSnowball: a statistical approach to extracting entity relationships, *Proceedings of the 18th international conference on World wide web,* 2009, pp. 101-110.
9. *Wu F., Weld D.S.* Open information extraction using Wikipedia, *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics,* 2010, pp. 118-127.

10. *Fader A., Soderland S., Etzioni O*. Identifying relations for open information extraction, *Proceedings of the Conference on Empirical Methods in Natural Language Processing,* 2011, pp. 1535-1545.
11. *Etzioni O., Fader A., Christensen J. et al.* Open information extraction: The second, *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence,* 2011, pp. 3-10.
12. *Schmitz M., Bart R., Soderland S. et al.* Open language learning for information extraction, *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning,* 2012, pp. 523-534.
13. *Angeli G., Johnson Premkumar M.J., Manning C.D.* Leveraging linguistic structure for open domain information extraction, *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing,* 2015, pp. 344-354.
14. *Nakashole N., Weikum G., Suchanek F.* PATTY: A taxonomy of relational patterns with semantic types, *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning,* 2012, pp. 1135-1145.
15. *Amurskaya O.Yu., Egorova A.D.* Sintaksicheskiy parsing kak analiz struktury predlozheniya (sintagmy) [Syntactic parsing as an analysis of sentence structure (syntagma)], *Filologiya i kul'tura* [Philology and Culture], 2022, No. 4 (70), pp. 14-23.
16. *Kobzareva T.Yu.* Lingvisticheskiy bazis analiza poverkhnostno-sintaksicheskikh svyazey segmentov v russkom predlozhenii [Linguistic basis for the analysis of superficial-syntactic connections of segments in a Russian sentence], *Vestnik RGGU. Seriya: Istoriya. Filologiya. Kul'turologiya. Vostokovedenie* [Bulletin of the Russian State University for the Humanities. Series: History. Philology. Cultural Studies. Oriental Studies], 2008, No. 6, pp. 157-170.
17. *Kobzareva T.Yu.* Ierarkhiya zadach poverkhnostno-sintaksicheskogo analiza russkogo predlozheniya [Hierarchy of tasks of superficial syntactic analysis of Russian sentences], *Nauchno-tekhnicheskaya informatsiya. Seriya 2: Informatsionnye protsessy i sistemy* [Scientific and technical information. Series 2: Information processes and systems], 2007, No. 1, pp. 23-35.
18. *Kobzareva T.Yu., Lakhuti D.G., Nozhov I.M.* Segmentatsiya russkogo predlozheniya: poverkhnostno-sintaksicheskiy analiz kak samostoyatel'nyy modul' analiza teksta [Segmentation of Russian sentence: surface-syntactic analysis as an independent module of text analysis], *Mater. 5-oy mezhdunarodnoy konferentsii «Informatsionnoe obshchestvo, informatsionnye resursy i tekhnologii telekommunikatsii». Sektsiya «Intellektual'nye sistemy avtomatizirovannoy podderzhki nauchnykh issledovaniy»* [Proceedings of the 5th international conference "Information society, information resources and telecommunication technologies". Section "Intelligent systems of automated support of scientific research"]. Moscow: VINITI. NTI, 2000, pp. 31-34.
19. *Kobzareva T.Yu.* Problema koreferentsii v ramkakh poverkhnostno-sintaksicheskogo analiza russkogo yazyka [he problem of coreference in the framework of superficial syntactic analysis of the Russian language], *Komp'yuternaya lingvistika i intellektual'nye tekhnologii: Tr. Mezhdunarodnoy konferentsii «Dialog 2003»* [Computer linguistics and intellectual technologies: Proceedings of the International Conference "Dialogue 2003"]. Moscow: Nauka, 2003, pp. 278 -284.
20. *Kobzareva T.Yu.* Printsipy segmentatsionnogo analiza russkogo predlozheniya [Principles of segmentation analysis of Russian sentences], *Moskovskiy lingvisticheskiy zhurnal* [Moscow Linguistic Journal], 2004, Vol. 8, No. 1, pp. 31-80.
21. *Kobzareva T.Yu.* Poisk khozyaina predlozhnoy gruppy v russkom predlozhenii [Search for the owner of a prepositional group in a Russian sentence], *Komp'yuternaya lingvistika i intellektual'nye tekhnologii: Tr. Mezhdunarodnoy konferentsii «Dialog-2010»* [Computer linguistics and intellectual technologies: Proceedings of the International Conference "Dialogue-2010"], 2010, Issue 9 (16), pp. 186-191.
22. *Rets I., Astruc L., Coughlan T., Stickler U.* Approaches to simplifying academic texts in English: English teachers' views and practices, *English for Specific Purposes,* 2022, Issue 68, pp. 31-46.
23. *Crossley S.A.*, *Allen D., McNamara D.S.* Text simplification and comprehensible input: A case for an intuitive approach, *Language Teaching Research,* 2012, Issue 16, No. 1, pp. 89-108. DOI: 10.1177/1362168811423456.
24. *Soemer A., Schiefele U.* Text difficulty, topic interest, and mind wandering during reading, *Learning and Instruction,* 2019, Issue 61, No. 1, pp. 12-22.
25. *Allen D.* A study of the role of relative clauses in the simplification of news texts for learners of English, *System,* 2009, Issue 37, No. 4, pp. 585-599.
26. *Long M.H.* Optimal input for language learning: Genuine, simplified, elaborated, or modified elaborated?, *Language Teaching,* 2020, Issue 53, No. 2, pp. 169-182. DOI: 10.1017/S0261444819000466.

27. *Tickoo M.L*. Simplification: Theory and Application. Anthology Series 31, *ERIC Clearinghouse,* 1993, pp. 254.
28. Big-O, *Big-O.io*. Available at: https://big-o.io (accessed 03 July 2024).
29. Complexity Cheat Sheet for Python Operations, *GeeksforGeeks*. Available at: https://www.geeksforgeeks.org/complexity-cheat-sheet-for-python-operations/ (accessed 03 July 2024).

Статью рекомендовал к опубликованию д.т.н. В.Ф. Лубенцов.

**Герасименко Евгения Михайловна** – Южный федеральный университет; e-mail: egerasimenko@sfedu.ru; г. Таганрог, Россия; тел.: 88634371651; кафедра систем автоматизированного проектирования им В.М. Курейчика; доцент.

**Кравченко Юрий Алексеевич –** e-mail: yakravchenko@sfedu.ru; тел.: 88634371651; кафедра систем автоматизированного проектирования им В.М. Курейчика; д.т.н.; профессор.

**Шаненко Дарья Андреевна –** e-mail: shanenko@sfedu.ru; тел.: 88634371651; кафедра систем автоматизированного проектирования им В.М. Курейчика, ассистент.

**Gerasimenko Evgeniya Mihailovna –** Southern Federal University; e-mail: egerasimenko@sfedu.ru; Taganrog, Russia; phone: +78634371651; the Department of Computer Aided Design named after V.M. Kureychik; associate professor.

**Kravchenko Yury Alekseevich –** e-mail: yakravchenko@sfedu.ru; phone: +78634371651; the Department of Computer Aided Design named after V.M. Kureychik; dr. of eng. sc.; professor.

**Shanenko Daria Andreevna –** e-mail: shanenko@sfedu.ru; phone: +78634371651; the Department of Computer Aided Design named after V.M. Kureychik; assistant.

УДК 004.056.55 DOI 10.18522/2311-3103-2024-5-102-118

**С.В. Поликарпов, В.А. Прудников, К.Е. Румянцев**

## СИНТЕЗ ПСЕВДО-ДИНАМИЧЕСКИХ ФУНКЦИЙ PD-sbox-ARX-32

*Целью работы является разработка метода синтеза оптимальных псевдо-динамических функций PD-sbox-ARX-32, размерностью 32 бита, в соответствие с противоречивыми требованиями к криптографическим характеристикам, рассматриваемой структуры. Рассмотрены методы синтеза классических sbox, в том числе с использованием эволюционного и генетического методов. Представлены требования к криптографическим характеристикам, как к функциям PD-sbox, так и к их составным элементам (классические sbox и ARX-функции). Предложен метод синтеза псевдо-динамических функций PD-sbox-ARX-32, включающий два этапа: 1) эвристический поиск структуры, соответствующей противоречивым требованиям к результирующим криптографическим характеристикам, потребляемым программным и аппаратным ресурсам, а также скорости работы представленной функции; 2) поиск оптимальных параметров основного элемента PD-sbox-ARX-32 – ARX-функций, при помощи эволюционного метода, суть которого заключается в подборе значений циклических сдвигов в ARX-функциях. В результате получен набор четырёх ARX-функций для псевдо-динамического преобразования PD-sbox-ARX-32, имеющего вес линейных характеристик равный $2^{-13}$ и разностных характеристик равный $2^{-32}$ (при этом эмпирический вес составляет $2^{-26}$). Для определения весов криптографических характеристик в работе применены методы на основе использования SAT-решателей. Приведены выводы о том, что подобранная структура 32-битной ARX-функции в составе PD-sbox позволяет обеспечить критический путь (максимальное количество последовательных операций сложения по модулю $2^{16}$) в четыре раза меньше чем 8-итерационная 32-битная Alzette-подобная структура, при двухкратном увеличении количества операций и при сопоставимых максимальных значениях весов разностных и линейных характеристик. Аналогичный результат получается при сравнении 32-битной ARX-функции с 8-итерационным 32-битным преобразованием из блочного криптоалгоритма Speck32. Предложенный метод синтеза параметров 32-битной ARX-функции позволяет минимизировать количество затрачиваемых ассемблерных инструкций на операции циклического сдвига при реализации на малоресурсных 8-битных микроконтроллерах AVR (например, ATmega328P).*

*Криптографические свойства; псевдо-динамическая функция; PD-sbox-ARX-32; синтез псевдо-динамических функций.*

**S.V. Polikarpov, V.A. Prudnikov, K.E. Rumyantsev**

## SYNTHESIS OF PSEUDO-DYNAMIC FUNCTIONS PD-sbox-ARX-32

*The aim of the work is to develop a method for synthesizing optimal pseudo-dynamic functions PD-sbox-ARX-32, 32-bit in size, in accordance with conflicting requirements for cryptographic characteristics of the considered structure. The methods for synthesizing classical sbox'es are considered, including those using evolutionary and genetic methods. The requirements for cryptographic characteristics are presented, both for the PD-sbox functions and for their constituent elements (classical sbox and ARX functions). A method for synthesizing pseudo-dynamic functions PD-sbox-ARX-32 is proposed, including two stages: 1) heuristic search for a structure corresponding to conflicting requirements for the resulting cryptographic characteristics, consumed software and hardware resources, as well as the speed of operation of the presented function; 2) search for optimal parameters of the main element of PD-sbox-ARX-32 – ARX functions, using the evolutionary method, the essence of which is to select the values of cyclic shifts in ARX functions. As a result, a set of four ARX functions was obtained for the pseudo-dynamic transformation of PD-sbox-ARX-32, having the weight of linear characteristics equal to* $2^{-13}$ *and difference characteristics equal to* $2^{-32}$ *(in this case the empirical weight is* $2^{-26}$*). To determine the weights of cryptographic characteristics, methods based on the use of SAT solvers were used in the work. The paper concludes that the selected structure of the 32-bit ARX function in the PD-sbox allows for a critical path (maximum number of sequential addition operations modulo* $2^{16}$*) that is four times smaller than that of the 8-iteration 32-bit Alzette-like structure, with a twofold increase in the number of operations and comparable maximum values of the weights of the difference and linear characteristics. A similar result is obtained when comparing the 32-bit ARX function with the 8-iteration 32-bit transformation from the Speck32 block cryptographic algorithm. The proposed method for synthesizing the parameters of the 32-bit ARX function allows for minimizing the number of assembler instructions spent on cyclic shift operations when implemented on low-resource 8-bit microcontrollers AVR (for example ATmega328P).*

*Cryptographic properties; pseudo-dynamic function; PD-sbox-ARX-32; synthesis of pseudo-dynamic functions.*

**Введение.** Преобразование sbox (s-box) является сокращением от substitution box – блок подстановки/замены (узел замены, в терминологии ГОСТ 28147-89), под термином «подстановка» подразумевается операция в виде табличной замены значений, операция не обязательно должна быть взаимно однозначной, как, например, в криптоалгоритме DES [1]. Функция sbox является основным нелинейным элементом множества блочных криптоалгоритмов, обеспечивающим противодействие к различным видам криптоанализа. sbox обладает множеством параметров, напрямую влияющих на устойчивость, как самой функции, так и полноценного криптоалгоритма, в котором sbox применён. Следовательно, при синтезе этого нелинейного элемента должно учитываться множество противоречивых требований, связанных со стойкостью sbox к различным криптографическим атакам, потреблению программных и аппаратных ресурсов, а также его быстродействию при программной или аппаратной реализации. Функция sbox и различные её разновидности обладают следующими характеристиками: линейные, разностные, алгебраические, а также их производные. Одним из направлений развития sbox является их объединение в структуру псевдо-динамической функции PD-sbox, основной особенностью которой является возможность разрушения статистических взаимосвязей между входными и выходными значениями за счёт динамической трансформации их криптографических свойств, а также параллельная работа фиксированных нелинейных элементов, входящих в состав PD-sbox [2]. Псевдо-динамические функции, семейства PD-sbox-ARX являются перспективным направлением развития псевдо-динамических преобразований PD-sbox, являющихся основным нелинейным элементом псевдослучайной функции pCollapser. Отличие PD-sbox-ARX от PD-sbox заключается в использовании ARX-функций в составе псевдо-динамических преобразований в качестве фиксированного нелинейного элемента [3]. Стоит отметить, что применение в качестве фиксированных sbox линейных ARX-функции показало, что миниверсия типовой функции на основе SP-сети не способна обеспечить хоть какой-то уровень нелинейности. Однако, миниверсия псевдослучайной функции

pCollapser позволяет получить из набора 4 ARX-конструкций, обладающих крайне низкими криптографическими свойствами, качественную нелинейную функцию. Подобное проявление свойств pCollapser подтверждает правильность концепции псевдо-динамических подстановок PD-sbox (разрушение статистических взаимосвязей между входными и выходными значениями за счёт динамической трансформации их криптографических свойств) [4].

**Цель работы** – определение метода синтеза оптимальных псевдо-динамических функций PD-sbox-ARX-32, размерностью 32 бита, в соответствие с противоречивыми требованиями к криптографическим свойствам, затрачиваемым ресурсам и задержки преобразования рассматриваемой структуры.

**Описание псевдо-динамических функций PD-sbox.** Псевдо-динамическая функция (преобразование) PD-sbox представляет собой структуру, включающую в свой состав набор фиксированных sbox или иных нелинейных элементов, в частности – ARX-функций [5, 6].

Аргумент каждого фиксированного sbox параметризован значением состояния $S_i$, где $i$ – номер фиксированного sbox или иного нелинейного элемента (от 0 до $N-1$). Текущее значение состояния $S = \{S_0, S_1, S_2, \dots, S_{N-1}\}$ задаёт один вариант преобразования из набора возможных PD-sbox. Преобразование, полученное на основе определённого значения состояния $S$ следует называть эквивалентным (сгенерированным) преобразованием для PD-sbox. Число эквивалентных преобразований ограничено количеством возможных значений состояния $S$, которые могут динамически изменяться в ходе обработки блока информации. При этом предполагается, что вероятностные свойства состояния $S$ соответствуют равномерному распределению.

Общий вид выражения, описывающего структуру псевдо-динамической функции PD-sbox:

$$\mathrm{Y} = \bigoplus_{\mathrm{i}=0}^{\mathrm{N}-1} \mathrm{sbox_i}(\mathrm{X} \oplus \mathrm{S_i}),$$

где sbox – фиксированный блок замены (обычно описывается в виде табличной замены значений, при этом операция не обязательно должна быть взаимнооднозначной, как, например, в криптоалгоритме DES [1]); $N$– количество фиксированных sbox; $X$ – биты на входе; $Y$ – биты на выходе; $S$ – биты состояния псевдо-динамической функции; $\oplus$ – операция сложения по модулю 2.

Псевдо-динамическая функция PD-sbox может функционировать в статическом (ключезависимом) и динамическом (зависимом от значений ключа и промежуточных состояний). В случае динамического равновероятного изменения состояний $S$, как дифференциальные усреднённые свойства, так и линейные, могут быть близки к идеальным (при усреднении характеристик по всем эквивалентным преобразованиям). Это потенциально позволяет нейтрализовать существующие методы дифференциального и линейного криптоанализа [5, 6].

**Описание псевдо-динамических функций PD-sbox-ARX.** В [3] предложен вариант применения специально подобранных ARX-функций в составе псевдо-динамических операций sbox, что позволяет обеспечить как параллелизм обработки информации, так и стойкость к статистическим методам криптоанализа и возможность эффективной программной реализации. Структура используемых ARX-функций приведена на рис. 1. Выбор подобной архитектуры функции обусловлен обеспечением криптографических свойств и оптимальным использованием возможностей современных процессоров и аппаратных платформ.

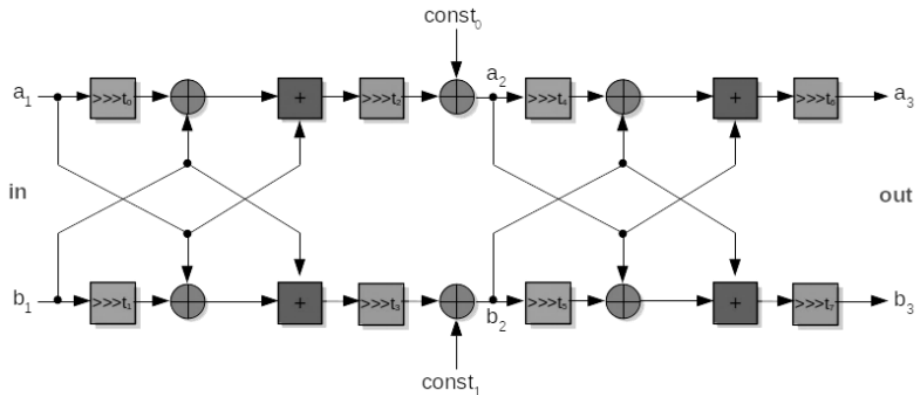



*Рис. 1. Структура используемых ARX-функций*

PD-sbox-ARX состоит из четырёх параллельно включённых в её структуру ARX-функций. Размерность входа-выхода PD-sbox-ARX соответствует размерности используемых ARX-конструкций. На рис. 2 представлена псевдо-динамическая sbox, включающая в свой состав четыре параллельно интегрированных ARX-функции. Размерность входа-выхода PD-sbox-ARX соответствует размерности используемых ARX-конструкций [2].

Выражение, описывающее значение на выходе:

$$c_i = \bigoplus_{j=0}^{3} funcARX_j(m_i \oplus s_j^i),$$

где $i$ – индекс-битного слова из входного/выходного вектора и далее индекс PD-sbox-ARX; $j$ – индекс компонента PD-sbox-ARX; $m_i$ – $n$-битные слова из входного вектора; $c_i$ – $n$-битные слова из выходного вектора; $funcARX$ – ARX-функция; $s_j^i$ – $n$-битные слова из входного вектора управляющего состояния [3].

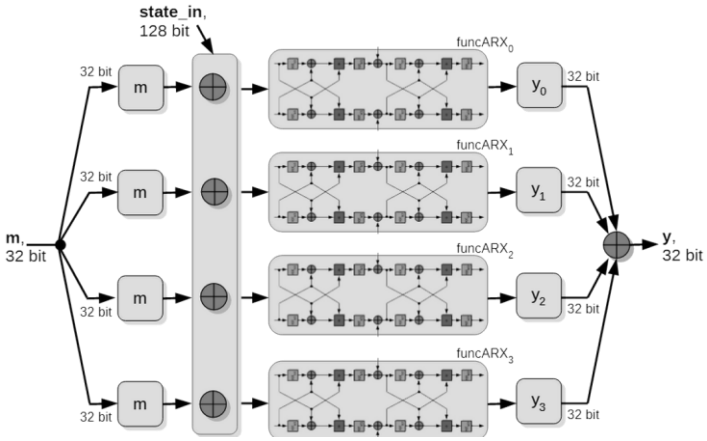



*Рис. 2. Псевдо-динамическая операция sbox на основе ARX-конструкций*

Выражение, описывающее индивидуальные управляющие состояния на выходе PD-sbox-ARX:

$$g_n^i = c_i \oplus funcARX_j\left(m_i \oplus s_j^i\right) = \bigoplus_{n=0, n \neq i}^{3} funcARX_j(m_i \oplus s_j^i).$$

**Описание sbox Alzette.** В работе [7] представлен sbox размерностью 64 бит, реализованный на основе ARX-конструкций – Alzette. Данное преобразование может быть вычислено с использованием 12 инструкций на современных процессорах, а параллельная реализация sbox Alzette может использовать векторные (SIMD) инструкции. За одну итерацию Alzette достигает разностных и линейных свойств, сравнимых со свойствами sbox AES [7]. Alzette обладает следующим преимуществом – использование операций размерностью 32 бит, следовательно, согласно [8], его эффективная реализация возможна на множестве различных архитектур, благодаря использованию регистров сдвига (barrel shift registers), если они доступны, а также благодаря подобранным значениям операций циклического сдвига.

Рассмотрим структуру Alzette, представленную на рис. 3. Функция параметризована константой $c$ размерностью 32 бит, используемой 4 раза. На вход Alzette поступает два слова размерностью 32 бит. Далее над каждым из них выполняются операции циклического сдвига, сложения по модулю $2^{32}$, а также XOR.

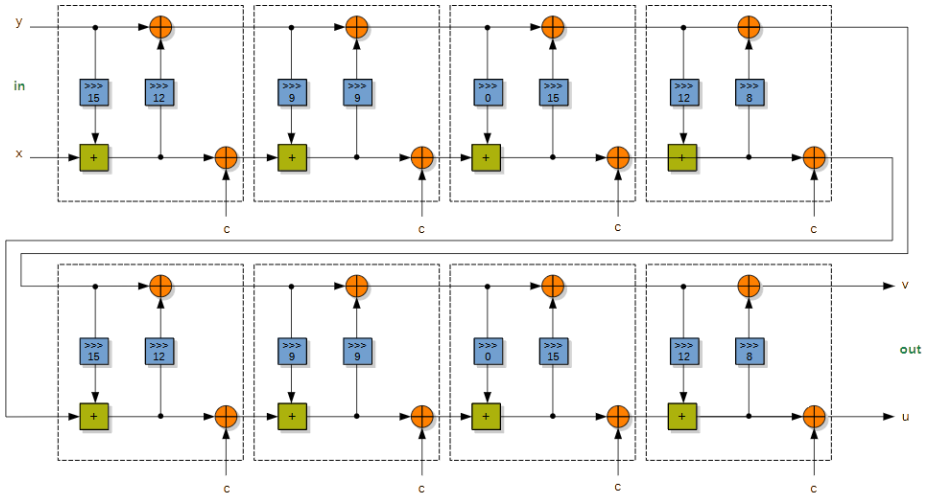



*Рис. 3. Структура sbox Alzette (8 раундов)*

При использовании Alzette в составе преобразования Sparkle, используются следующие константы: $c_0 = b7e15162$, $c_1 = bf715880$, $c_2 = 38b4da56$, $c_3 = 324e7738$, $c_4 = bb1185eb$, $c_5 = 4f7c7b57$, $c_6 = cfbfa1c8$, $c_7 = c2b3293d$. Sparkle – это семейство криптографических преобразований размерностью 256, 384 или 512 бит. Sparkle включает в свой состав операции сложения по модулю слова, циклического сдвига и XOR (ARX-конструкции). Основой перестановки являются sbox на основе ARX-функций Alzette [9].

Таким образом, применение ARX-операций позволило разработчикам Alzette создать функцию (блок замены) размерностью 64 бит. Это значительно больше, чем типовая размерность блоков замены в криптографических преобразованиях. Следовательно, прямой расчёт криптографических свойств (например, таблиц распределения разностей DDT и линейных приближений LAT) для sbox Alzette крайне затруднителен или вовсе невозможен, так как требует недоступного объёма вычислительных ресурсов. Поэтому, авторы Alzette применяли SAT-решатели для определения её криптографических свойств [7].

Применительно к sbox Alzette её авторы используют термин “раунд” (итерация), по аналогии с блочными криптографическими преобразованиями (например, как в криптоалгоритме Speck64, структура которого представлена на рис. 4). В табл. 1 приведены границы разностных и линейных свойств для версий Alzette с использованием от 1 до 12 раундов. Основная версия Alzette предполагает 4 раунда преобразования [10].

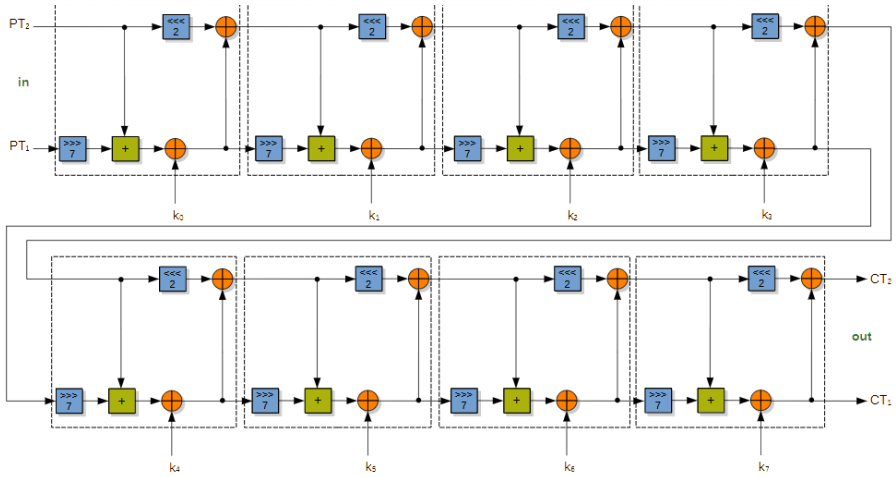



*Рис. 4. Структура Speck32 (8 раундов)*

Таблица 1

**Нижние границы для разностных и линейных свойств sbox Alzette**

| $(r_0, r_1, r_2, r_3, s_0, s_1, s_2, s_3)$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| (31, 17, 0, 24, 24, 17, 31, 16) | 0<br>0 | 1<br>0 | 2<br>1 | 6<br>2 | 10<br>5 | 18<br>8 | ≥24<br>13(11.64) | ≥32<br>17(15.79) | ≥36<br>– | ≥42<br>– | ≥46<br>– | ≥52<br>– |
| (17, 0, 24, 31, 17, 31, 16, 24) | 0<br>0 | 1<br>0 | 2<br>1 | 6<br>2 | 10<br>5 | 17<br>9 | ≥25<br>13 | ≥31<br>16 | ≥37<br>– | ≥41<br>– | ≥47<br>– | –<br>– |
| (0, 24, 31, 17, 31, 16, 24, 17) | 0<br>0 | 1<br>0 | 2<br>1 | 6<br>2 | 10<br>6 | 18<br>8 | ≥24<br>13 | ≥32<br>15 | ≥36<br>– | ≥42<br>– | –<br>– | –<br>– |
| (24, 31, 17, 0, 16, 24, 17, 31) | 0<br>0 | 1<br>0 | 2<br>1 | 6<br>2 | 10<br>5 | 17<br>9 | ≥25<br>12 | ≥31<br>16 | ≥37<br>– | –<br>– | –<br>– | –<br>– |
| Speck64 | 0<br>0 | 1<br>0 | 3<br>1 | 6<br>3 | 10<br>6 | 15<br>9 | 21<br>13 | 29<br>17 | ≥32<br>19 | –<br>21 | –<br>24 | –<br>27 |

Для каждого смещения первая строка демонстрирует $\log_2 p$, где $p$ – максимальная ожидаемая вероятность дифференциального следа для дифференциального случая. Вторая строка демонстрирует $\log_2 c$, где $c$ – максимальная ожидаемая абсолютная линейная корреляция следа для линейного случая. Значение, указанное в скобках, соответствует максимальной абсолютной корреляции линейной оболочки с учетом кластеризации, полученной экспериментальной проверкой.

Как и в блочных преобразованиях, авторы Alzette используют широко распространённый метод поиска разностных и линейных характеристик – когда определяются разностные и линейные свойства отдельных раундов и результирующие свойства всего преобразования определяются через «принцип накопления» (piling-up principle) [11]. Данный принцип предполагает, что входные значения каждого раунда являются статистически независимыми.

Стоит отметить одну важную особенность – в Alzette между «раундами» не используется добавление раундовых ключей, что, предположительно, существенно нарушает «принцип накопления». Однако, это не помешало авторам Alzette получить валидные криптографические свойства используя SAT-решатели и применяя метод Мацуи для поитерационного поиска разностных и линейных характеристик Alzette [10].

**Анализ криптографических свойств PD-sbox-ARX-32.** Рассмотрим анализ криптографических характеристик синтезированной псевдо-динамической функции PD-sbox-ARX-32, в частности разностных и линейных, а также сравним их с аналогичными параметрами миниверсии sbox Alzette, размерностью 32 бита.

В силу размерности и сложности анализируемых конструкций, для поиска разностных и линейных свойств использован метод поиска криптографических характеристик, использующий SAT-решатели, в частности – фреймворк CASCADA [12]. Создатели дан-

ного фреймворка особое внимание уделили анализу ARX-функций. В настоящее время SAT-решатели являются одним из основных методов поиска и валидации криптографических свойств и характеристик криптографических преобразований [12–15].

Свойства, используемые в разностном криптоанализе, являются разностями $(\alpha, \beta)$ по функции шифрования $E_k$ с высокой ожидаемой разностной вероятностью. При наличии разностей $(\alpha, \beta)$ по $f$ его разностная вероятность определяется как

$$\#\{x: f(x\Delta\,\alpha)\nabla f(x) = \beta\}/2^n\,,$$

где $\Delta = \oplus = \nabla$. Оператор $\nabla$ вычисляет разность пары значений $(x, x')$, а оператор $\Delta$ принимает в качестве входных данных значение $x$ и разность $\alpha$ и выводит такое значение $x'$, что пара $(x, x')$ имеет разность $\alpha$.

Ожидаемая разностная вероятность $p$ — это разностная вероятность, усредненная по ключевому пространству $K$:

$$p = \frac{1}{|K|}\sum_{k\in K}\#\{x: f(x\Delta\alpha)\nabla\mathrm{f(x)} = \beta\}/2^n,$$

а сложность разностного криптоанализа составляет $O(1/p)$.

Свойство $(\alpha, \beta)$ над функцией $f$ является действительным, если его вероятность распространения не равна нулю. В этом случае мы определяем вес распространения $(\alpha, \beta)$ как отрицательный двоичный логарифм его вероятности распространения:

$$PW_f(\alpha, \beta) = -\log_2\Big(PP_f(\alpha, \beta)\Big).$$

Результат работы CASCADA – веса найденных оптимальных криптографических характеристик:

$$w = -\log_2 P(.),$$

где $P(.)$ – вероятность появления входной/выходной разности для тестируемой функции (для разностного анализа) или значения корреляции (для линейного анализа).

Следующие обозначения и определения являются стандартными в линейном криптоанализе.

**Определение 1.** Итеративный блочный шифр – это алгоритм, который преобразует блок открытого текста фиксированного размера $n$ в блок шифр-текста идентичного размера под воздействием ключа $k$ путем применения итеративного обратимого преобразования $p$, называемого раундовым преобразованием. Обозначая открытый текст как $x_0$, а шифр-текст как $x_R$, операция зашифрования может быть записана как:

$$x_{r+1} = p_{k_r}(x_r), \quad r = 1, 2, \dots, R,$$

где $k_r$ являются подключами, сгенерированными алгоритмом выработки ключей. Для простоты мы рассматриваем -битные ключи и подключи.

**Определение 2.** Пусть $F: \{0,1\}^n \to \{0,1\}^n$ – биективное преобразование, $a, b$ – две маски $\in \{0,1\}^n$. Если $X \in \{0,1\}^n$ равномерно распределенная случайная величина, тогда смещение линейной аппроксимации $LB(a, b)$ определено, как:

$$LB(a, b) = \Pr_X\{a * X = b * F(X)\} - \frac{1}{2},$$

где $*$ – скалярное произведение. Если $F$ параметризовано ключом $K$, запишем $LB(a, b, K)$ и ожидаемое линейное отклонение $ELB(a, b)$ определено, как:

$$ELB(a, b) = \underset{K}{E}\big(LB(a, b, K)\big).$$

Линейное отклонение может быть вычислено для различных преобразований, например, для одного *sbox*, раундовой функции или блочного шифра. Точное определение линейного отклонения является вычислительно сложной задачей по мере увеличения размерности преобразования.

**Определение 3.** Одноraундовая характеристика для раунда $i$ итерационного блочного шифра представляет собой пару-битных векторов $\langle a_i, b_i\rangle$, соответствующих входным и выходным маскам для этого раунда. -раундовая характеристика для раундов $1 \dots R$ представляет собой $(R + 1)$-кортеж $n$-битных векторов $\Omega = \langle a_1, a_2, \dots, a_{R+1}\rangle$, где $\langle a_i, a_{i+1}\rangle$ соответствуют входным и выходным маскам для раунда $i$.

**Определение 4.** Шифр Маркова – блочный шифр, в котором линейные (и разностные) отклонения разных раундов независимы друг от друга, предполагая, что в разных раундах используются равномерно-случайные подключи [16].

В нашем случае PD-sbox-ARX является небиективным преобразованием, однако, возможным вариантом его применения является использование в качестве нелинейной функции Фейстель-подобного итерационного преобразования, представленного на рис. 5. Которое, как известно, является взаимооднозначным преобразованием. Возможность применения небиективного преобразования является важным преимуществом сети Фейстеля и такие известные криптоалгоритмы как DES и ГОСТ [1, 17] используют небиективную функцию. Стоит отметить, что изначально понятия итерационной характеристики было введено для анализа криптоалгоритма DES [18, 19].

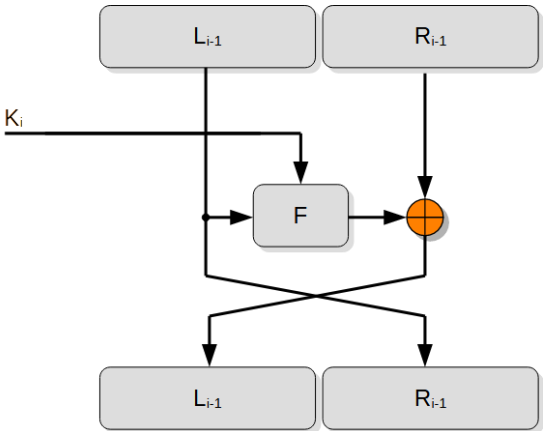



*Рис. 5. Раунд сети Фейстеля*

PD-sbox-ARX-32 является невзаимнооднозначной псевдо-динамической функцией, несмотря на это, её анализ с использованием SAT/SMT-решателей возможен и найденные характеристики будут валидны. В качестве примера следует привести работу [20], в которой представлены результаты первого криптоанализа шифра DES с использованием SAT-решателей.

**Особенности программной реализации на малоресурсных процессорах.** Реализация криптографических преобразований на микроконтроллерах (малоресурсных процессорах) является важнейшей задачей, обусловленной как повсеместным использованием криптографических преобразований в сетевых протоколах, так и значительной ресурсоёмкости этих преобразований, что может привести к дефициту вычислительных ресурсов для основных процессов.

Стоит отметить, что на различных процессорах/микроконтроллерах для ARX-операций может требоваться разное количество тактов на их выполнение. Однако, для относительно простых микроконтроллеров предполагается, что на выполнение операций ADD или XOR требуется один такт [21]. Но выполнение операции ROL может требовать значительного количества инструкций и тактов микроконтроллера.

В качестве таких процессоров/микроконтроллеров мы будем рассматривать широко распространённые 8-битные микроконтроллеры AVR фирмы Atmel, микроконтроллеры ATmega328P (применяется, например, в Arduino UNO R3) и микроконтроллеры на базе инструкций MIPS32/MIPS64 (самым известным является семейство микроконтроллеров PIC32 от Microchip).

Мы не будем рассматривать здесь более совершенные процессоры/микроконтроллеры на основе архитектур RISC-V и ARM, в которых уже реализована встроенная операция ROL, требующая небольшого количества тактов на выполнение. Что касается 8-битных микроконтроллеров семейства AVR (например, ATmega328P), то в них аппаратно реализована инструкция циклического сдвига только на 1 бит.

Для определения количества инструкций на операции циклического сдвига с разным количеством сдвигаемых бит мы воспользовались ресурсом godbolt.org, интерфейс которого представлен на рис. 6, позволяющим в интерактивном режиме, как вывести результат компиляции исходного кода в виде набора ассемблерных инструкций, так и легко выбрать компилятор и архитектуру/семейство целевого процессора или микроконтроллера.

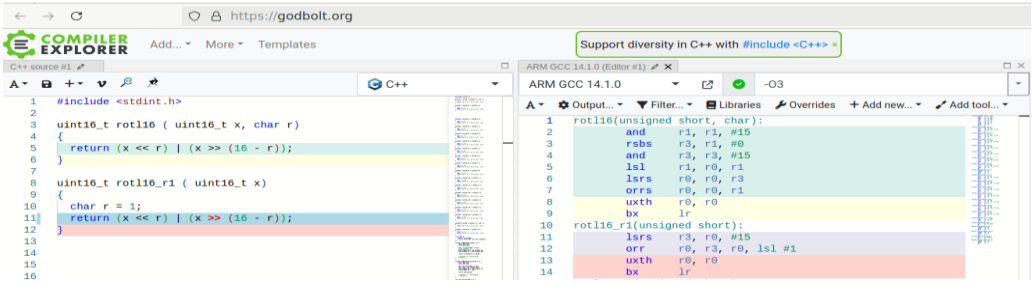



*Рис. 6. Результат компиляции на ресурсе godbolt.org*

Мы взяли типовой способ описания операции ROL на языке C в виде типовой конструкции из двух нециклических сдвигов (влево и вправо) и объединения результатов при помощи операции OR или XOR. Как известно, современные компиляторы, в том числе gcc, хорошо распознают такую типовую конструкцию и при компиляции заменяют её на соответствующую встроенную команду. Или, в случае отсутствия такой команды – на оптимальный набор инструкций, выполняющий эквивалентное преобразование.

Для каждого значения циклического сдвига мы реализовали отдельную функцию, представленную на рис. 7, что позволило в дальнейшем оценить количество затрачиваемых инструкций для ROL с разным значением сдвига.

```
1   uint16_t rotl16_r1 (uint16_t x) {char r = 1; return (x << r) | (x >> (16 - r));}
2   uint16_t rotl16_r2 (uint16_t x) {char r = 2; return (x << r) | (x >> (16 - r));}
3   uint16_t rotl16_r3 (uint16_t x) {char r = 3; return (x << r) | (x >> (16 - r));}
4   ...
5   uint16_t rotl16_r15 (uint16_t x) {char r = 15; return (x << r) | (x >> (16 - r));}
```

*Рис. 7. Отдельные функции для значений циклического сдвига*

После компиляции для каждой функции подсчитали количество требуемых на реализацию ассемблерных инструкций. При этом инструкцию RET (возврат из функции) мы не учитывали, так как при компиляции целиком всей ARX-функции обычно осуществляется встраивание кода ROL-функции непосредственно в точку вызова функции (вместо её фактического вызова, что позволяет снизить количество инструкций вызова и возврата). В табл. 2 приведён пример результата компиляции, использовался флаг «-O3».

Таблица 2

**Пример результата компиляции 16-битовых операций ROL-1 и ROL-2 для различных целевых платформ**

| Операция | Архитектура / компилятор | |
|---|---|---|
| | x86-64 / gcc 14.2 | AVR / gcc 14.1.0 |
| ROL-1 | rotl16 r1(unsigned short):<br>mov eax, edi<br>rol ax<br>ret | rotl16 r1(unsigned int):<br>.L__stack_usage = 0<br>lsl r24<br>rol r25<br>adc r24,__zero_reg__<br>ret |

*Окончание табл. 2*

| | | |
|---|---|---|
| ROL-2 | rotl16 r2(unsigned short):<br>mov eax, edi<br>rol ax, 2<br>ret | rotl16 r2(unsigned int):<br>.L__stack_usage = 0<br>mov r18,r24<br>mov r24,r25<br>mov r19,r18<br>swap r19<br>lsr r19<br>lsr r19<br>andi r19,lo8(3)<br>lsl r25<br>lsl r25<br>swap r24<br>lsr r24<br>lsr r24<br>andi r24,lo8(3)<br>lsl r18<br>lsl r18<br>or r24,r18<br>or r25,r19<br>ret |

Как можно увидеть, для архитектуры x86-64 требуется всего 2 инструкции для реализации ROL-1 и ROL-2, а вот для AVR требуется 3 инструкции для ROL-1 и 17 инструкций для ROL-2. При этом для x86-64 одна из инструкций – команда перемещения из одного регистра в другой (mov eax, edi), которая может быть убрана при оптимизации кода.

В табл. 3 приведены сводные значения по количеству инструкций на реализацию операции 16-битового циклического сдвига для рассматриваемых архитектур.

Таблица 3

**Количество ассемблерных инструкций для реализации 16-битовой операций ROL в зависимости от значения циклического сдвига**

| **Операция** | **Количество инструкций для архитектуры/компилятора** | | | | |
|---|---|---|---|---|---|
| | AVR (gcc 14.1.0) | Arduino Uno (1.8.9) | mips32/64 (gcc 14.1.0) | ARM (gcc 14.1.0) | x86-64 (gcc 14.2) |
| ROL-1 | **3** | **3** | 5 | 3 | 2 |
| ROL-2 | 17 | 18 | | | |
| ROL-3 | 17 | 18 | | | |
| ROL-4 | 13 | 14 | | | |
| ROL-5 | 17 | 18 | | | |
| ROL-6 | 17 | 18 | | | |
| ROL-7 | 13 | 14 | | | |
| ROL-8 | **3** | **3** | | | |
| ROL-9 | 13 | 14 | | | |
| ROL-10 | 17 | 18 | | | |
| ROL-11 | 17 | 18 | | | |
| ROL-12 | 13 | 14 | | | |
| ROL-13 | 17 | 18 | | | |
| ROL-14 | 17 | 18 | | | |
| ROL-15 | **4** | **4** | | | |

Следует обратить внимание на то, что иные операции (сложение по модулю, XOR) соответствуют одной ассемблерной инструкции и ими можно пренебречь.

**Метод синтеза псевдо-динамической функции PD-sbox-ARX-32.** В ходе экспериментальных исследований нами был выработан следующий метод синтеза PD-sbox-ARX-32:

1. Эвристический выбор структуры ARX-функции с учётом возможных особенностей программной и аппаратной реализаций.

2. Начальное заполнение параметров циклических сдвигов ARX-функций (всего по 8 значений на 4 функции) значением 8.

3. Последовательный выбор каждого параметра ARX-функции, замена его на случайное значение из допустимого диапазона (от 0 до 15), проверка криптографических свойств (вес разностных и линейных характеристик) и ожидаемого количества затрачиваемых ассемблерных инструкций для полученной версии ARX-функции. После обхода всех параметров первой ARX-функции выбирается наилучшая версия (при её наличии), которая заменяет исходную. Далее осуществляется переход к следующей ARX-функции для выполнения аналогичных действий.

4. Действия из пункта 3 повторяются до момента отсутствия улучшений в свойствах ARX-функций.

5. Формирование при помощи пунктов 2 и 3 набора наиболее удачных ARX-функций.

6. Выбор из набора наиболее удачных ARX-функций варианта с наименьшим количеством затрачиваемых ассемблерных инструкций для микроконтроллеров архитектуры AVR. В табл. 4 представлено количество ассемблерных инструкций для реализации 16-битовых операций ROL для 12 отобранных параметров PD-sbox-ARX-32.

Следует обратить внимание на следующее: по пункту 2 экспериментальные исследования показали, что если сразу задать «удачный» вариант значений циклического сдвига (для 16-битных сдвигов это значение равно 8), то значительно увеличивается вероятность того, что эти значения будут в результирующей ARX-функции после операций синтеза; по пункту 5 экспериментальные исследования показали, что предложенный пошаговый подбор параметров позволяет получать PD-sbox-ARX с достаточно близкими криптографическими свойствами. При синтезе 100 PD-sbox-ARX 73 варианта имели вес разностных характеристик $Wd$ равный 32 и вес линейных характеристик $Wl$, равный 13 и 14. Такие характеристики очень близки 8-раундовым преобразованиям Speck32 и miniAlzette32. Поэтому варианты с более худшими характеристиками исключаются из набора.

В результате получилось 12 параметров, представленных в табл. 4, при этом наименьшее количество ассемблерных инструкций для архитектуры AVR составило $N = 360$, что в 1,4 раза и 1,7 раза больше, чем для наихудшего и наилучшего вариантов синтеза соответственно, для которых $N = 256$ и $N = 217$.

Таблица 4

**Количество ассемблерных инструкций для реализации 16-битовых операций ROL для 12 отобранных параметров PD-sbox-ARX-32**

| № | Архитектура микроконтроллера | | |
|---|---|---|---|
| | AVR | mips64 | ARM |
| 1 | 242 | 150 | 73 |
| 2 | 248 | | |
| 3 | 256 | | |
| 4 | 222 | | |
| 5 | 242 | | |
| 6 | 248 | | |

*Окончание табл. 4*

| | | | |
|---|---|---|---|
| 7 | 240 | | |
| 8 | 231 | | |
| 9 | **217** | | |
| 10 | 234 | | |
| 11 | 234 | | |
| 12 | 248 | | |

В табл. 5 приведено сравнение свойств лучшей синтезированной PD-sbox-ARX-32 со свойствами 8-итерационной 32-битной Alzette-подобной структуры и 8-итерационным 32-битным преобразованием из блочного криптоалгоритма Speck32.

Таблица 5

**Количество ассемблерных инструкций для реализации 16-битовой операций ROL**

| **Преобразование** | **Архитектура микроконтроллера** | | | **Криптографические свойства** | | | |
|---|---|---|---|---|---|---|---|
| | AVR | mips64 | ARM | $Wd$ | $Wde$ | $Wl$ | $Wle$ |
| miniAlzette32 (8 раундов) | 154 | 70 | 42 | 27 | ~27 | 13 | ~13 |
| Speck32 (8 раундов) | 240 | 80 | 48 | 24 | ~24 | 12 | ~12 |
| **PD-sbox-ARX-32** | **217** | **150** | **73** | **32** | **~26** | **13** | **~13** |

**Сравнение разработанного метода синтеза с методом случайного поиска параметров псевдо-динамической функции PD-sbox-ARX-32.** Для оценки эффективности предложенного метода нами сформировано 100 000 случайных наборов параметров ARX-функций для PD-sbox-ARX-32, для которых определено количество затрачиваемых ассемблерных инструкций и криптографические свойства – вес разностных характеристик $Wd$, вес линейных характеристик $Wl$. Данный универсальный метод применён нами для сравнения, так как иные методы синтеза параметров PD-sbox-ARX не представлены в открытой печати.

Ниже приведены результаты в виде гистограмм распределения по количеству затрачиваемых ассемблерных инструкций – на рис. 8, в виде гистограмм распределения по весам разностных $Wd$ и линейных $Wl$ характеристик – на рис. 9.

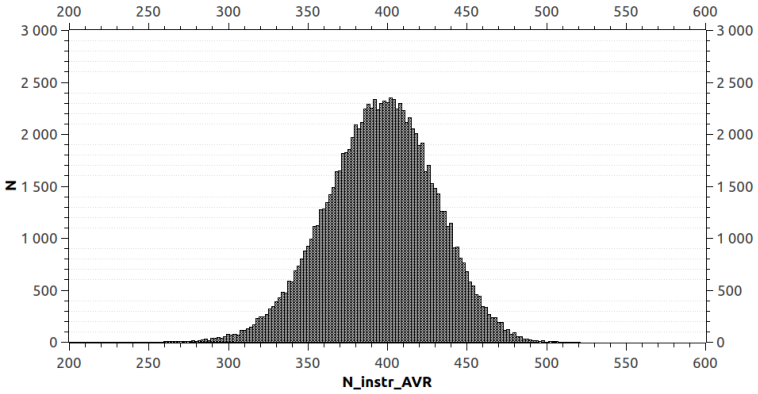



*Рис. 8. Гистограмма распределения наборов параметров по количеству затрачиваемых ассемблерных инструкций*

Полученная гистограмма имеет нормальное распределение, среднее значение составляет 395,92, стандартное отклонение 33,96, стандартная ошибка 0,1074, минимальное 230, максимальное 512, медиана 397.

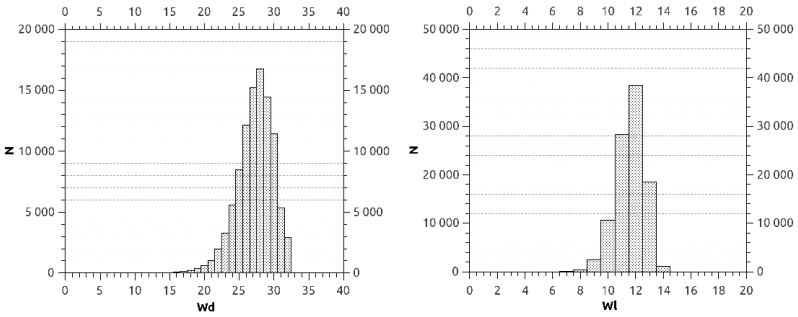



*Рис. 9. Гистограмма распределения по весам разностных $Wd$ и линейных $Wl$ характеристик*

Анализ гистограммы, представленной на рис. 8, позволяет выделить 5 комбинаций параметров, выделенных на рис. 10, обладающих минимальным количеством затрачиваемых ресурсов. Свойства параметров приведены в табл. 6.

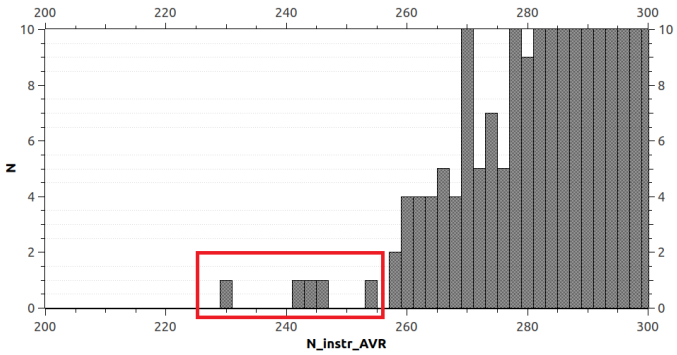



*Рис. 10. Комбинации параметров, обладающих минимальным количеством затрачиваемых ресурсов при использовании метода случайного подбора*

Таблица 6

**Свойства 5 случайно подобранных комбинаций параметров**

| N | AVR | $Wd$ | $Wl$ |
|---|---|---|---|
| 1 | 230 | 24 | 9 |
| 2 | 243 | 20 | 10 |
| 3 | 244 | 18 | 8 |
| 4 | 245 | 20 | 11 |
| 5 | 246 | 25 | 12 |
| **Синтезированный PD-sbox-ARX** | **217** | **32** | **13** |

Все случайно подобранные комбинации параметров ARX-функций с минимальным количеством затрачиваемых ассемблерных инструкций обладают неприемлемыми криптографическими характеристиками. Однако, даже в данном случае они существенно уступают синтезированному PD-sbox-ARX по количеству затрачиваемых ресурсов. При сравнении с вариантом 1 из табл. 6, разница составляет ~5%, учитывая его неудовлетворительные криптографические характеристики.

На рис. 11 приведены результаты в виде гистограмм распределения по количеству затрачиваемых ассемблерных инструкций, минимальное значение затрачиваемых ассемблерных инструкций равно 284 при $Wd > 29$ и $Wl > 10$.

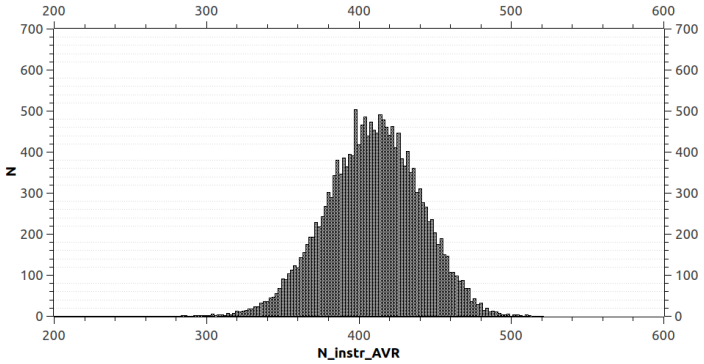



*Рис. 11. Гистограмма распределения наборов параметров по количеству затрачиваемых ассемблерных инструкций при $Wd > 29$ и $Wl > 10$*

Полученная гистограмма имеет нормальное распределение, среднее значение составляет 409,66, стандартное отклонение 0,57, стандартная ошибка 0,2264, минимальное 284, максимальное 553, медиана 410.

Анализ гистограммы, представленной на рис. 11, позволяет выделить 5 комбинаций параметров, выделенных на рис. 12, обладающих минимальным количеством затрачиваемых ресурсов при $Wd > 29$ и $Wl > 10$. Свойства параметров приведены в табл. 7.

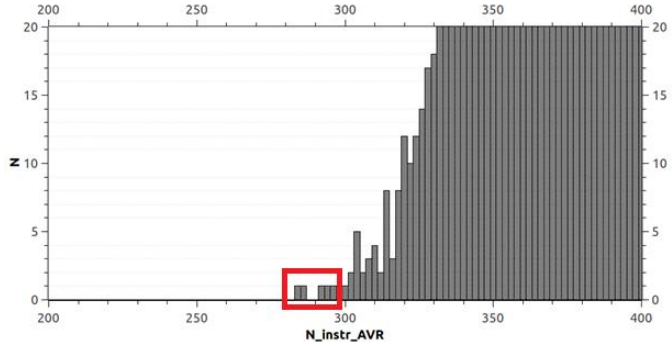



*Рис. 12. Комбинации параметров, обладающих минимальным количеством затрачиваемых ресурсов при использовании метода случайного подбора при $Wd > 29$ и $Wl > 10$*

Таблица 7

**Свойства 5 случайно подобранных комбинаций параметров при $Wd > 29$ и $Wl > 10$**

| N | AVR | $Wd$ | $Wl$ |
|---|---|---|---|
| 1 | 284 | 30 | 11 |
| 2 | 286 | 30 | 11 |
| 3 | 292 | 30 | 11 |
| 4 | 294 | 31 | 11 |
| 5 | 297 | 30 | 11 |
| **Синтезированный PD-sbox-ARX** | **217** | **32** | **13** |

Все случайно подобранные комбинации параметров ARX-функций с минимальным количеством затрачиваемых ассемблерных инструкций обладают удовлетворимыми криптографическими свойствами, однако существенно уступают синтезированному PD-sbox-ARX по количеству затрачиваемых ресурсов. При сравнении с вариантом 1 из таблицы 7, разница составляет ~24%, без учёта уступающих криптографических свойств.

**Заключение.** Подобранная структура 32-битной ARX-функции в составе PD-sbox позволяет обеспечить критический путь (максимальное количество последовательных операций сложения по модулю $2^{16}$) в четыре раза меньше, чем 8-итерационная 32-битная Alzette-подобная структура, при двухкратном увеличении количества операций и при сопоставимых максимальных значениях весов разностных и линейных характеристик.

Аналогичный результат получается при сравнении 32-битной ARX-функции с 8-итерационным 32-битным преобразованием из блочного криптоалгоритма Speck32.

**Практическая значимость:** при аппаратной реализации ARX-функции данное свойство позволяет пропорционально уменьшить (до 4 раз) задержку при преобразовании блоков информации.

Предложенный метод синтеза параметров 32-битной ARX-функции позволяет получить параметры операций циклического сдвига, при которых обеспечивается максимальный вес разностной характеристики равный $2^{-32}$ (эмпирический вес $2^{-26}$) и вес линейной характеристики $2^{-13}$ для результирующего PD-sbox-ARX, включающей в свой состав четыре 32-битные ARX-функции. Сопоставимые разностные и линейные характеристики имеют 8-итерационные 32-битная Alzette-подобная структура и 8-итерационное 32-битное преобразование из блочного криптоалгоритма Speck32.

Предложенный метод синтеза параметров 32-битной ARX-функции позволяет минимизировать количество затрачиваемых ассемблерных инструкций на операции циклического сдвига при реализации на малоресурсных 8-битных микроконтроллерах семейства AVR (например, ATmega328P).

Например, при параметрах циклических сдвигов [5, 8, 8, 8, 8, 11, 0, 0], [12, 3, 8, 8, 8, 15, 4, 4], [8, 12, 8, 8, 9, 8, 8, 8], [1, 8, 8, 8, 8, 3, 12, 12] на их реализацию потребуется 217 ассемблерных инструкций микроконтроллера AVR. Что ориентировочно на 10% меньше, чем для типовых получаемых параметров и соответствует количеству ресурсов 8-итерационного преобразования из криптоалгоритма Speck32 (240 ассемблерных инструкций микроконтроллера AVR) на циклические сдвиги.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Diffie W. and Hellman M.E.* Special Feature Exhaustive Cryptanalysis of the NBS Data Encryption Standard // in Computer. – June 1977. – Vol. 10, No. 6. – P. 74-84. – DOI: 10.1109/C-M.1977.217750.
2. *Поликарпов С.В., Румянцев К.Е., Кожевников А.А.* Псевдо-динамические таблицы подстановки: основа современных симметричных криптоалгоритмов // Научное обозрение. – 2014. – № 12. – С. 162–166. – URL: http://www.sced.ru/ru/files/7_12_1_2014/7_12_1_2014.pdf.
3. *Поликарпов С.В., Румянцев К.Е., Прудников В.А.* Высокопроизводительная псевдослучайная функция pCollapserARX256-32x2 // XXIV научно-практическая конференция «РусКрипто'2022». – 2022. – URL: https://ruscrypto.ru/resource/archive/rc2022/files/02_polikarpov_rumyantsev_prudnikov.pdf.
4. *Поликарпов С.В., Румянцев К.Е., Прудников В.А.* Исследование свойств миниверсии псевдослучайной функции pcollapser // Известия ЮФУ. Технические науки. – 2022. – № 6 (230). – С. 148-162. – DOI: 10.18522/2311-3103-2022-6-148-162.
5. *Поликарпов С.В., Румянцев К.Е., Кожевников А.А.* Исследование линейных характеристик псевдо-динамических подстановок // Известия ЮФУ. Технические науки – 2015. – № 5 (166). – С. 111-123. – URL: http://izv-tn.tti.sfedu.ru/wp-content/uploads/2015/5/11.pdf.
6. *Polikarpov S., Rumyantsev K., Petrov D.* Computationally efficient method for determining averaged distribution of differentials for pseudo-dynamic substitutions // International Conference on Electrical, Electronics, Materials and Applied Science, AIP Conf. Proc., 1952 / eds. V. Rao, A. Ben, S. Bhukya. Amer. Inst. Phys., 2018, UNSP 020091. – DOI: 10.1063/1.5032053.
7. *Beierle C. et al.* Alzette: A 64-Bit ARX-box / In: Micciancio, D., Ristenpart, T. (eds) //Advances in Cryptology – CRYPTO 2020: Lecture Notes in Computer Science. – 2020. – Vol. 12172. – Springer, Cham. – https://doi.org/10.1007/978-3-030-56877-1_15.

8. *Dinu D., Corre Y.L., Khovratovich D. et al*. Triathlon of lightweight block ciphers for the Internet of things // J Cryptogr Eng. – 2019. – 9. – P. 283-302. – https://doi.org/10.1007/s13389-018-0193-x.
9. *Beierle C., Biryukov A., Cardoso dos Santos L., Großschädl J., Perrin L., Udovenko A., Velichkov V., & Wang Q.* Lightweight AEAD and Hashing using the Sparkle Permutation Family // IACR Transactions on Symmetric Cryptology. – 2020. – 2020(S1). – P. 208-261. – https://doi.org/10.13154/tosc.v2020.iS1.208-261.
10. *Beierle C., Biryukov A., Cardoso dos Santos L.* Schwaemm and Esch: Lightweight Authenticated Encryption and Hashing using the Sparkle Permutation Family. University of Luxembourg. – 2019. – URL: https://sparkle-lwc.github.io/assets/sparkle-specification-latest.pdf.
11. *Beyne T.* A Geometric Approach to Linear Cryptanalysis / In: Tibouchi, M., Wang, H. (eds) // Advances in Cryptology – ASIACRYPT 2021: Lecture Notes in Computer Science. – Vol. 13090. – Springer, Cham, 2021. – https://doi.org/10.1007/978-3-030-92062-3_2.
12. *Ranea A., Rijmen V.* Characteristic Automated Search of Cryptographic Algorithms for Distinguishing Attacks (CASCADA) // IET Information Security. – 2022. – 16 (6). – DOI: 10.1049/ise2.12077. – URL: https://eprint.iacr.org/2022/513.pdf.
13. *Stachowiak S., Kurkowski M., & Soboń A*. New results in SAT – cryptanalysis of the AES // 2022 IEEE 16th International Scientific Conference on Informatics (Informatics). – 2022. – P. 280-286.
14. *Bellini E., Piccoli A.D., Formenti M., Gérault D., Huynh P., Pelizzola S., Polese S., & Visconti A.* Differential Cryptanalysis with SAT, SMT, MILP, and CP: A Detailed Comparison for Bit-Oriented Primitives // Cryptology and Network Security. – 2023.
15. *Shi J., Liu G., & Li C.* SAT-Based Security Evaluation for WARP against Linear Cryptanalysis // IET Information Security. – 2023.
16. *Collard Baudoin & Standaert François-Xavier.* Experimenting linear cryptanalysis // Cryptology and Information Security Series. – 2011. – 7. – 10.3233/978-1-60750-844-1-1.
17. ГОСТ 28147-89 Системы обработки информации. Защита криптографическая. Алгоритм криптографического преобразования. – М.: Стандартинформ, 1990.
18. *Matsui Mitsuru.* Linear Cryptoanalysis Method for DES Cipher // Advances in Cryptology - EUROCRYPT '93, Workshop on the Theory and Application of of Cryptographic Techniques, Lofthus, Norway, May 23-27, 1993, Proceedings. – 1993. – P. 386-397. – DOI: http://dx.doi.org/10.1007/3-540-48285-7_33.
19. *Biham E., & Shamir A.* Differential Cryptanalysis of the Data Encryption Standard. – Springer: New York, 1993.
20. *Massacci F., Marraro L.* Logical cryptanalysis as a SAT-problem: Encoding and analysis // In Journal of Automated Reasoning. – 2000. – 24. – P. 165-203.
21. 8-bit Atmel Microcontroller with 128Kbytes In-System Programmable Flash // ATmega128, ATmega128L. Rev. 2467X–AVR–06/11. 2011 Atmel Corporation. – URL: http://ww1.microchip.com/ downloads/en/devicedoc/doc2467.pdf.

REFERENCES

1. *Diffie W. and Hellman M.E*. Special Feature Exhaustive Cryptanalysis of the NBS Data Encryption Standard, *in Computer,* June 1977, Vol. 10, No. 6, pp. 74-84. DOI: 10.1109/C-M.1977.217750.
2. *Polikarpov S.V., Rumyantsev K.E., Kozhevnikov A.A.* Psevdo-dinamicheskie tablitsy podstanovki: osnova sovremennykh simmetrichnykh kriptoalgoritmov [Pseudo-dynamic substitution tables: the basis of modern symmetric cryptoalgorithms], *Nauchnoe obozrenie* [Scientific Review], 2014, No. 12, pp. 162-166. Available at: http://www.sced.ru/ru/files/7_12_1_2014/7_12_1_2014.pdf.
3. *Polikarpov S.V., Rumyantsev K.E., Prudnikov V.A.* Vysokoproizvoditel'naya psevdosluchaynaya funktsiya pCollapserARX256-32x2 [High-performance pseudorandom function pCollapserARX256-32x2], *XXIV nauchno-prakticheskaya konferentsiya «RusKripto'2022»* [XXIV scientific and practical conference "RusCrypto'2022"], 2022. Available at: https://ruscrypto.ru/resource/archive/rc2022/files/02_polikarpov_rumyantsev_prudnikov.pdf.
4. *Polikarpov S.V., Rumyantsev K.E., Prudnikov V.A.* Issledovanie svoystv miniversii psevdo-sluchaynoy funktsii pcollapser [Study of properties of miniversion of pseudo-random function pcollapser], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2022, No. 6 (230), pp. 148-162. DOI: 10.18522/2311-3103-2022-6-148-162.
5. *Polikarpov S.V., Rumyantsev K.E., Kozhevnikov A.A.* Issledovanie lineynykh kharakteristik psevdo-dinamicheskikh podstanovok [Study of linear characteristics of pseudo-dynamic substitutions], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2015, No. 5 (166), pp. 111-123. Available at: http://izv-tn.tti.sfedu.ru/wp-content/uploads/2015/5/11.pdf.
6. *Polikarpov S., Rumyantsev K., Petrov D.* Computationally efficient method for determining averaged distribution of differentials for pseudo-dynamic substitutions, *International Conference on Electrical, Electronics, Materials and Applied Science, AIP Conf. Proc., 1952*, eds. V. Rao, A. Ben, S. Bhukya. Amer. Inst. Phys., 2018, UNSP 020091. DOI: 10.1063/1.5032053.

7. *Beierle C. et al.* Alzette: A 64-Bit ARX-box, In: Micciancio, D., Ristenpart, T. (eds), *Advances in Cryptology – CRYPTO 2020: Lecture Notes in Computer Science,* 2020, Vol. 12172. Springer, Cham. Available at: https://doi.org/10.1007/978-3-030-56877-1_15.
8. *Dinu D., Corre Y.L., Khovratovich D. et al*. Triathlon of lightweight block ciphers for the Internet of things, *J Cryptogr Eng*., 2019, 9, pp. 283-302. Available at: https://doi.org/10.1007/s13389-018-0193-x.
9. *Beierle C., Biryukov A., Cardoso dos Santos L., Großschädl J., Perrin L., Udovenko A., Velichkov V., & Wang Q.* Lightweight AEAD and Hashing using the Sparkle Permutation Family, *IACR Transactions on Symmetric Cryptology,* 2020, 2020(S1), pp. 208-261. Available at: https://doi.org/10.13154/tosc.v2020.iS1.208-261.
10. *Beierle C., Biryukov A., Cardoso dos Santos L.* Schwaemm and Esch: Lightweight Authenticated Encryption and Hashing using the Sparkle Permutation Family. University of Luxembourg, 2019. Available at: https://sparkle-lwc.github.io/assets/sparkle-specification-latest.pdf.
11. *Beyne T.* A Geometric Approach to Linear Cryptanalysis, In: Tibouchi, M., Wang, H. (eds), *Advances in Cryptology – ASIACRYPT 2021: Lecture Notes in Computer Science,* Vol. 13090. Springer, Cham, 2021. Available at: https://doi.org/10.1007/978-3-030-92062-3_2.
12. *Ranea A., Rijmen V.* Characteristic Automated Search of Cryptographic Algorithms for Distinguishing Attacks (CASCADA), *IET Information Security,* 2022, 16 (6). DOI: 10.1049/ise2.12077. Available at: https://eprint.iacr.org/2022/513.pdf.
13. *Stachowiak S., Kurkowski M., & Soboń A.* New results in SAT – cryptanalysis of the AES, *2022 IEEE 16th International Scientific Conference on Informatics (Informatics),* 2022, pp. 280-286.
14. *Bellini E., Piccoli A.D., Formenti M., Gérault D., Huynh P., Pelizzola S., Polese S., & Visconti A.* Differential Cryptanalysis with SAT, SMT, MILP, and CP: A Detailed Comparison for Bit-Oriented Primitives, *Cryptology and Network Security,* 2023.
15. *Shi J., Liu G., & Li C.* SAT-Based Security Evaluation for WARP against Linear Cryptanalysis, *IET Information Security,* 2023.
16. *Collard Baudoin & Standaert François-Xavier.* Experimenting linear cryptanalysis, *Cryptology and Information Security Series,* 2011, 7. 10.3233/978-1-60750-844-1-1.
17. GOST 28147-89 Sistemy obrabotki informatsii. Zashchita kriptograficheskaya. Algoritm kriptograficheskogo preobrazovaniya [GOST 28147-89 Information processing systems. Cryptographic protection. Cryptographic transformation algorithm]. Moscow: Standartinform, 1990.
18. *Matsui Mitsuru.* Linear Cryptoanalysis Method for DES Cipher, *Advances in Cryptology - EUROCRYPT '93, Workshop on the Theory and Application of of Cryptographic Techniques, Lofthus, Norway, May 23-27, 1993, Proceedings,* 1993, pp. 386-397. DOI: http://dx.doi.org/10.1007/3-540-48285-7_33.
19. *Biham E., & Shamir A.* Differential Cryptanalysis of the Data Encryption Standard. Springer: New York, 1993.
20. *Massacci F., Marraro L.* Logical cryptanalysis as a SAT-problem: Encoding and analysis, *In Journal of Automated Reasoning,* 2000, 24, pp. 165-203.
21. 8-bit Atmel Microcontroller with 128Kbytes In-System Programmable Flash, *ATmega128, ATmega128L. Rev. 2467X–AVR–06/11. 2011 Atmel Corporation*. Available at: http://ww1.microchip.com/ downloads/en/devicedoc/doc2467.pdf.

Статью рекомендовал к опубликованию д.т.н., профессор О.И. Шелухин.

**Поликарпов Сергей Витальевич** – Южный федеральный университет; e-mail: polikarpovsv@sfedu.ru, г. Таганрог, Россия; тел.: +79085159762; к.т.н.

**Прудников Вадим Александрович** – e-mail: pruvad@yandex.ru, тел.: +79198961427; старший преподаватель.

**Румянцев Константин Евгеньевич** – e-mail: rke2004@mail.ru; тел.: +79281827209; д.т.н.; профессор.

**Polikarpov Sergey Vitalievich** – Southern Federal University; e-mail: polikarpovsv@sfedu.ru; Taganrog, Russia; phone: +79085159762; cand. of eng. sc.

**Prudnikov Vadim Aleksandrovich** – e-mail: pruvad@yandex.ru; phone: +7 9198961427; senior lecturer.

**Rumyantsev Konstantin Evgenyevich** – e-mail: rke2004@mail.ru; phone: +79281827209; dr. of eng. sc.; professor.

УДК 004.932 DOI 10.18522/2311-3103-2024-5-119-131

**А.А. Александров, Г.С. Мизюков, М.А. Бутакова**

# ОЦЕНКА КАЧЕСТВА СЛИЯНИЯ ИЗОБРАЖЕНИЙ С ИСПОЛЬЗОВАНИЕМ ЭНТРОПИИ ШЕННОНА И КОЭФФИЦИЕНТА ПОЛЕЗНОЙ ИНФОРМАЦИИ ХАРТЛИ

*Исследуются методы повышения качества изображений, получаемых из гетерогенных источников информации на основе многомодальной интеграции. Дополнительная информация из нескольких модальностей позволяет использовать признаки, которые невозможно правильно интерпретировать, если анализировать информацию отдельно от одного источника. В качестве подтверждения актуальности темы рассматриваются современные исследования в данной области. Целью работы является повышение информативности изображений, получаемых в результате слияния данных от разнородных источников, и получение высококачественных изображений, пригодных для корректной работы алгоритмов машинного обучения. Для достижения поставленной цели авторами решается ряд задач: создание подхода к измерению качества изображений, в рамках которого необходимо разработать ряд алгоритмов, описывающих процесс оценки качества результатов слияния на основе многомодальной информации; реализация полученных алгоритмов в программной среде для валидации предлагаемого подхода; проведены оценочных экспериментов на основе представленных алгоритмов, в частности, вычисления мер информативности изображений и влияния шумов и размытости на энтропию объединённого изображения. Результаты экспериментальных исследований на наборах данных из открытых источников показали, что предложенный метод позволяет определить наилучший вариант слияния изображений, при котором данные будут иметь максимальную информативность. Использование энтропии по Шеннону дает возможность вычислить количество информации, передаваемой в изображениях, а коэффициент полезной информации Хартли позволяет оценить количество присутствующих шумов в изображении. Также, в статье проводится сравнение результатов при различных уровнях шума и степени размытости изображений, демонстрирующее различные результаты алгоритмов при оценке качества изображений. Предложенный подход проиллюстрирован на примере анализа изображений, полученных путём слияния данных от двух типов приборов – инфракрасной камеры и видеокамеры, снимающей изображения в видимом диапазоне.*

*Слияние данных; многомодальная информация; энтропия информации; коэффициент Хартли; оценка качества изображений.*

**A.A. Alexandrov, G.S. Miziukov, M.A. Butakova**

# IMAGE FUSION QUALITY ASSESSMENT USING SHANNON ENTROPY AND HARTLEY USEFUL INFORMATION COEFFICIENT

*The article examines methods for improving the quality of images obtained from different sources of information through multi-modal integration. By combining information from various modalities, features that cannot be accurately interpreted when analyzed separately can be utilized. To support the relevance of this topic, recent research in this area is discussed. The goal of the work is to enhance the information content of images resulting from merging data from diverse sources and create high-quality images suitable for accurate machine learning applications. To achieve this objective, the authors address several tasks. They develop an approach for measuring image quality and design algorithms to evaluate the quality of fused results based on multi-modal information. These algorithms are implemented within a software framework for validating the proposed approach. Evaluation experiments are conducted based on the presented calculations of the information content of images and the effect of noise and blur on the entropy of the combined image. The results of the experimental studies on data sets from open sources have shown that the proposed method allows determining the best way to merge images with maximum information content. The use of Shannon entropy makes it possible to calculate the amount of information transmitted in images, and the Hartley coefficient of useful information helps estimate the amount of noise in an image. Additionally, the article compares the results at different levels of noise and degrees of blur in images, demonstrating the effectiveness of different algorithms for evaluating image quality. To illustrate the proposed approach, we analyze images obtained by combining data from two devices: an infrared camera and a video camera capturing images in the visible range.*

*Data fusion; multimodal information; information entropy; Hartley coefficient; image quality assessment.*

**Введение.** Актуальность оценки качества изображения обусловлена непрерывным развитием технологий визуализации и обработки изображений. С растущим распространением технологий машинного обучения и искусственного интеллекта, применяемых в задачах распознавания объектов, возрастает потребность в больших объемах качественных изображений. Изображения представляют собой визуальные объекты, содержащие информацию, а в системах распознавания они рассматриваются как наборы пикселей с различными цветовыми характеристиками. Для того, чтобы научить систему распознавать на изображениях объекты, необходимо предоставить ей большое количество примеров с четко обозначенными границами объектов и качественными входными данными. Изображения должны иметь высокое разрешение, с минимальным количеством шумов, искажений и содержать интересующий объект. После оценки точности и полноты распознавания объектов результаты работы модели иногда оказываются неудовлетворительными, в этом случае возникает необходимость в дополнительном обучении на новых данных и/или использовании дополнительной информации. Однако существуют проблемы, связанные с определением качества изображений, поскольку отсутствуют строгие объективные критерии, а также восприятие качества зависит от субъективной оценки человека. Для повышения качества распознавания на сегодняшний день в различных системах применяют многомодальную интеграцию (объединение) данных, например: в медицине – для создания новых изображений с целью улучшения зрительного восприятия [1, 2]; в «умных» автомобилях, в которых используют данные от различных источников, таких как RGB-камера, Depth-камера, FIR-камера, LIDAR, другие сенсоры и датчики [3, 4]; в интеллектуальных системах контроля для повышения эффективности принятия решений [5]; в беспилотных аппаратах и спутниках, где слияние данных от различных источников дистанционного зондирования позволяет повысить точность классификации [6] и во многих других областях.

Многомодальная интеграция данных не только позволяет собрать большие объемы информации от различных источников, но и предусматривает более рациональное использование информации. Такой подход имеет важное значение в тех случаях, когда данные могут предоставить дополнительную или избыточную информацию, которая позволяет уменьшить неопределенность и повысить надежность систем компьютерного зрения. Предоставляемая дополнительная информация от источников различных типов позволяет использовать признаки, которые в рамках каждой отдельной модальности не представляют значимости. Имеется множество исследований по разработке методов объединения многомодальных данных, которые в англоязычной литературе получили название «Data Fusion» («Слияние данных»), например, [7–9]. Авторами данной статьи так же ранее проводилось исследование [5] по объединению информации от гетерогенных источников для повышения эффективности принятия решений при обнаружении дефектов на поверхности различных материалов. В этой работе был предложен комплексный метод объединения данных различной модальности для обогащения наборов признаков, характеризующих измеряемый объект при оценке наличия дефекта.

**Постановка задачи.** Одним из критериев оценки результатов слияния является достоверность полученной информации. Вопрос достоверности информации в системах слияния данных является важным критерием при оценке полученных результатов [10]. В многочисленных публикациях, например, в [3–6] сообщается об улучшении характеристик полученных данных с применением методов Data Fusion. Исследователи полагают, что полученные результаты безусловно будут более точные и содержать больше информации. Однако лишь небольшое число публикаций посвящено проблеме исследования достоверности полученных результатов. При анализе информации специалистам приходится самостоятельно определять достоверность полученных данных и надежность их источника. Мы предлагаем оценивать непосредственно исходные данные и результаты их слияния, выбирая наилучший вариант для дальнейшей работы с полученными изображениями.

При разработке метрик слияния изображений принято проводить принципиальное различие между субъективными и объективными оценками. Субъективная оценка (зрительное восприятие человека) надежна, трудоемка и дорога. Существует высокая по-

требность в объективных оценочных показателях для оценки качества объединенных изображений. В последние годы исследователями предложено множество показателей для объективной оценки качества изображений, таких как среднеквадратическая ошибка, пиковое отношение сигнал/шум и мера индекса структурного сходства [11]. Для оценки качества полученных объединённых медицинских изображений, при сравнении с эталонными снимками авторы в работе [12] применяют такие метрики для сравнения эталонных изображений и полученных в результате слияния данных, как уровень шума, четкость или размытость краев, контрастность, форму, структуру объектов и информационную энтропию – для характеристики текстуры изображения.

В теории информации основной мерой информации является энтропия, которая количественно определяет степень неопределенности, связанную с вероятностью появления тех или иных символов при передаче сообщений. Энтропия Шеннона оценивает среднее количество информации (математическое ожидание), которое содержится в значениях случайной величины появления тех или иных символов. Отождествляя энтропию с информацией, Клод Шеннон пришел к выводу, что количество информации, приобретаемое при полном выяснении состояния некоторой физической системы, равно энтропии этой системы. Энтропия обращается в ноль, когда одно из состояний достоверно, а остальные невозможны. При заданном числе состояний энтропия обращается в максимум, когда эти состояния равновероятны, а при увеличении числа состояний она увеличивается. Энтропии присуще свойство аддитивности, т.е. когда несколько независимых систем объединяются в одну, их энтропии суммируются. Энтропия Шеннона двух независимых событий $X$ и $Y$ равна сумме энтропий этих событий, согласно формуле (1).

$$H\left(X \cdot Y\right) = H\left(X\right) + H\left(Y\right). \quad (1)$$

Энтропия Шеннона используется для определения количества информации, содержащейся в сообщении, и минимальной длины сообщения, необходимой для передачи информации. Также она применяется для оценки информационного содержания случайной величины появления символа в сообщении, которая выражается в единицах информации, например в битах, и для определения оптимального способа сжатия данных без потерь. В работе мы рассматриваем изображения, для которых исследуется случайная величина появления пикселя с определённым цветом. Согласно формуле Хартли, вероятность события появления пикселя определенного цвета обратно пропорциональна числу исходов, если события равновероятны, то есть цвета распределены равномерно. Тогда вероятности появления пикселей каждого цвета будут равны. В случае, если преобладает определенный цвет, то вероятность его появления будет выше. Вероятностный подход Шеннона обобщает меру количества информации по Хартли, когда не все события являются равновероятностными, а также учитывает влияние шума в канале связи. В соответствии с формулой Шеннона информация, добавленная каждым новым символом, представляет собой логарифм от вероятности его появления, в основание которого ставится количество символов в алфавите. Аналогично сообщениям, состоящим из символов алфавита, изображения состоят из пикселей различных цветов. Добавление новых пикселей приводит к увеличению информации, содержащейся в изображении. Следовательно, информацию, добавленную каждым новым пикселем, можно расчитать как логарифм от вероятности появления пикселя определенного цвета, в основании которого стоит количество цветов в изображении. Эффективность метрик, использующих обобщение энтропии Шеннона, была продемонстрирована во многих исследованиях, где были показаны хорошие результаты [13–18]. Таким образом, информационная энтропия Шеннона может использоваться для оценки качества изображений.

Оценка качества изображения является актуальной темой для исследования в области цифровой обработки изображений. Например, авторы в статье [19] исследуют различные улучшенные модификации энтропии Шеннона для оценки информации о пространственном расположение пикселей в изображениях в градациях серого. Энтропию так же применяют в методах, позволяющих эффективно отражать информационное содержание инфракрасных изображений [20], для улучшения качества и надёжности кото-

рых авторы предлагают использовать оптимизированный генетический алгоритм и методы слияния изображений. Данный подход позволяет минимизировать шум, улучшить контраст и повысить пространственное разрешение, что приводит к улучшению качества инфракрасных изображений и обеспечивает более точный и надёжный количественный анализ. Другие исследователи применяют энтропию в методе оценки качества изображений без использования эталонов на основе слияния признаков, например, в работе [21] из изображений извлекаются признаки, в том числе и величина энтропии, затем на основе всех признаков формируется вектор признаков. Далее авторы с помощью алгоритмов машинного обучения получают зависимость между признаками изображения и показателями качества, что дает возможность оценивать качество самого изображения.

При слиянии двух и более изображений в одно, более информативное, могут появляться различного рода дефекты, такие как искажения, размытия, появление шумов. Также, при слиянии разных модальностей заранее не известно, какая из модальностей имеет большую информативность. Авторами предлагается использовать вычисление информативности изображений до и после объединения многомодальных данных, что позволяет оценить эффективность объединения данных, а также пригодность объединённого изображения для дальнейшего использования.

В данной работе, в разделе 3, приводится описание методов, разработанных авторами, для измерения информативности изображений. В разделе 4 предложен метод повышения информативности изображения на основе объединения многомодальных данных.

**Метод решения.** На качество изображения влияют в числе других такие параметры, как количество используемых цветов и глубина цвета, где каждому пикселю, формирующему изображение, присваивается информация о цвете. При работе с цифровыми изображениями человеческий глаз способен различить только определенный уровень глубины цвета, при его увеличении различия становятся почти незаметны. Во многом восприятие так же зависит от цветопередачи монитора, но системы работают непосредственно с пикселями, что позволяет использовать максимальное число цветов, которое может содержать изображение.

При осветлении теней на изображении или затемнении бликов скрытые различия в переходах между цветами могут проявиться в виде цветовых артефактов – отсутствия плавности переходов тонов и цветов изображения. Таким образом, увеличение контрастности будет приводить к уменьшению уровня глубины цвета изображения. Этот параметр влияет на плавность перехода тонов и цвета изображения. При увеличении глубины цвета соответственно увеличивается не только качество изображения, но и его объем. Для вычисления исходного объема изображения необходимо знать глубину цвета изображения. При определении глубины изображения воспользуемся формулой (2):

$$N = 2^i, \tag{2}$$

где $i$ – глубина изображения в битах на пиксель; $N$ – количество цветов в изображении.

Другим параметром качества изображения, характеризующим её информативность, является разрешение. Чем больше разрешение, тем больше мелких деталей может содержаться на изображении. Стоит учитывать, что при объединении данных изображения могут иметь различные разрешения. При приведении изображений к одним и тем разрешениям изображения масштабируются. При уменьшении размеров изображения происходит удаление пикселей, а при увеличении размеров с помощью методов интерполяции добавляются новые с учётом данных существующих пикселей. При этом объем изображения может сохраниться при использовании методов интерполяции с высокой точностью. Для определения объема изображения воспользуемся формулой (3).

$$V = \frac{(w \times h) \times i}{1024}, \tag{3}$$

где $w$ – ширина изображения в пикселях; $h$ – высота изображения в пикселях; $i$ – глубина изображения в байтах; $V$ – объема изображения в килобайтах.

На всех изображениях, полученных с помощью цифровой техники, присутствуют в определенной степени различные шумы. На цифровых изображениях шум может быть минимальным и не различимым, а быть максимально выраженным. Например, там, где на изображении должна быть однотонная и гладкая поверхность, присутствуют шероховатость или нечёткость. Особенную выраженность шум имеет на изображениях с наименьшей яркостью, так как яркие области имеют более сильный сигнал, следовательно, шум на ярких областях будет мало заметен. Помимо вариаций яркости, шум может быть хроматическим, при высокой концентрации которого, изображение может стать непригодным для обработки. Применение большого количества фильтров может привести к нежелательным ухудшениям, и изображение может приобрести характерные искажения, такие как потеря оттенков на переходах яркости и цвета, снижение насыщенности тонов, отображение в явном виде структуры растра. Для обнаружения шума в изображении воспользуемся формулой, известной как мера Хартли или хартлиевское количество информации. С помощью формулы (4) определим коэффициент полезной информации $I$ в изображении. Чем больше значение коэффициента, тем меньше шума в изображении.

$$I = \log_2 |S|, \tag{4}$$

где $S$ – множество всех возможных значений пикселей.

На рис. 1 приведены изображения, которые имеют различные уровни глубины цвета и различное количество шума по Гауссу. Исходное изображение с черно белым градиентом имеет разрешение 600×200 пикселей и глубину цвета 16 бит. К полученному изображению было добавлено небольшое количество шума по Гауссу 5% и 10% соответственно. Каждое изображение представлено в пяти вариантах глубины цвета: 16, 8, 4, 2 и 1 бит.

При слиянии изображений к одному изображению добавляется информация из второго. Очевидно, что каждое изображение является информативным, и при анализе из них можно извлечь все содержащиеся данные. Поэтому необходима мера, которая позволит количественно определить прирост информативности при слиянии двух изображений, а именно, как второе изображение влияет на информативность первого. Для этого воспользуемся определением энтропии по Шеннону, согласно формуле (5).

$$H = -\sum_{x=1}^{n} p(x) \log_2 p(x), \tag{5}$$

где $n$ – количество возможных событий; $p(x)$ – вероятность наступления события $x$, в нашем случае – вероятность появления того или иного цвета пикселя изображения.

С помощью формулы (5) определяем влияние наличия второго изображения на качество полученного изображения.

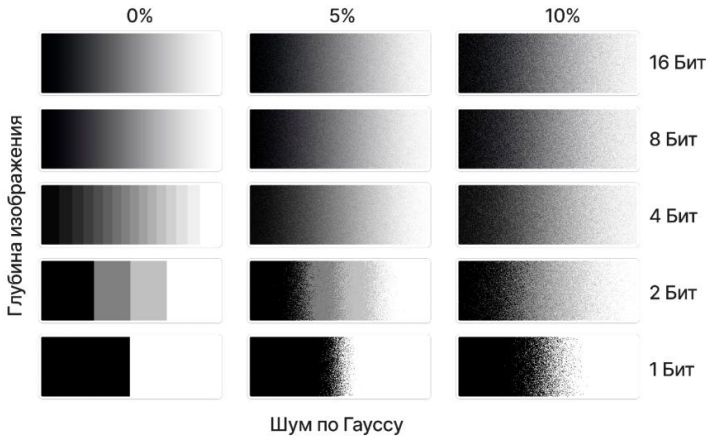



*Рис. 1. Изображения с различной глубиной цвета и различным количеством шума*

Чем больше коэффициент энтропии $H$, тем выше качество информации, передаваемой в изображении. Энтропия зависит только от вероятностей значений интенсивности, без учёта их пространственного распределения, что дает нам среднее количество информации о значениях интенсивности в изображении. В данном случае энтропия используется, как показатель степени сглаживания контуров, мелких деталей и краев изображения. Получается, что значение энтропии изображения, в котором есть размытые контуры, меньше по сравнению со значением энтропии исходного значения, следовательно коэффициент энтропии будет больше, если изображение будет с подчёркнутыми контурами рисунков.

Таким образом, имеется три формулы оценки качества изображения, такие как: глубина изображения, коэффициент полезной информации, коэффициент энтропии. Используя полученные выше формулы (2)–(5), можно установить, что чем больше значение критерия, тем меньше в изображении дефектов, следовательно выше его качество и информативность. Принятие взвешенного решения о качестве изображения можно представить в виде следующей формулы:

$$Q = \begin{cases} 1, \text{ если } \left(w_N \times N + w_I \times I + w_H \times H\right) \geq T_{\max}, \\ 0, \text{ если } T_{\min} \leq \left(w_N \times N + w_I \times I + w_H \times H\right) < T_{\max}, \\ -1, \text{ если } \left(w_N \times N + w_I \times I + w_H \times H\right) < T_{\min}, \end{cases} \quad (6)$$

где $w_N$, $w_I$ и $w_H$ – весовые коэффициенты параметров $N$, $I$, $H$, соответственно;

$T_{\min}$ и $T_{\max}$ – пороговые значения, которые определяют, при каком уровне качества изображение считается без дефектов или с дефектом.

Расчёт пороговых значений $T_{\min}$ и $T_{\max}$, является важным этапом при определении качества, необходимо установить оптимальные границы позволяющие классифицировать изображения по уровню качества. Расчёт проводится индивидуально для каждого набора данных, что позволяет учитывать специфику и особенности информации, содержащейся в изображениях. Первый способ основан на использовании мнения специалистов (экспертов) для определения пороговых значений. Эксперты анализируют изображения и предлагают свою субъективную оценку в зависимости от зрительного восприятия. Пороговые значения $T_{\min}$ и $T_{\max}$ будут определяться как среднее усеченное всех полученных значений. Второй способ, который мы применяем, использование выборки – из исходного множества данных случайным образом и алгоритмов кластеризации данных для разбиения полученных коэффициентов на заранее определённое число кластеров. Определив стабильное разбиение данных на кластеры и число кластеров, определяем пороговые значения $T_{\min}$ и $T_{\max}$, основываясь на анализе структуры кластеров.

Как видно из формулы (6), в случае если изображение нечитаемое, то значение мер будет минимальным, и результат вычисления будет меньше порогового значения $T_{\min}$. Результатом вычисления формулы (6) в этом случае будет «-1». Если значение коэффициента полезной информации в изображении низкое, что свидетельствует о наличии шума, и значение коэффициента энтропии $H$ уменьшается, что указывает на снижение сложности изображения, например изображение становится менее детализированным, то результат вычисления будет находиться между пороговыми значениями $T_{\min}$ и $T_{\max}$, свидетельствуя о среднем качестве изображения. В результате вычисления формулы (6) будет получен «0». Если результат вычисления превышает $T_{\max}$, это указывает на минимальное количество шума в изображении и максимальное количество содержимой информации. В результате вычисления формулы (6) будет получен «1». На основании полученных значений и формулы (6) можно делать выводы о качестве полученного изображения после слияния.

**Алгоритмы.** В этом разделе мы оценим качество реальных изображений с использованием формул для оценки качества изображения. Для демонстрации результатов подсчета информативности изображений воспользуемся алгоритмами, полученными из формул (2)–(6). Первым шагом является вычисление глубины изображения. Разработанная процедура определения глубины представлена в Алгоритме 1.

**Алгоритм 1:** Вычисление глубины изображения

```
Вход: путь к растровому изображению pathToImage
Результат: количество бит на пиксель
bitsPerPixel ← 0;
url ← new System.Uri(pathToImage);
source ← new BitmapImage(url); // Битовая карта изображения
bitsPerPixel ← source.Format.BitsPerPixel;
return bitsPerPixel;
```

Следующим этапом, зная значения глубины изображения, вычислим объем изображения. В Алгоритме 2 приведено описание вычисления информационного объёма изображения.

**Алгоритм 2:** Вычисление объема изображения

```
Вход: ширина изображения W; высота изображения H; глубина изображения в битах D
Результат: объем изображения в килобайтах
capacity ← 0;
capacity ← (W * H) * (N / 8); // Объем изображения в байтах
capacity ← capacity / 1024; // Перевод в более крупные единицы измерения информации
return capacity;
```

Наличие шума может понизить качество изображения маскируя тонкие детали изображения делая их размытыми и менее четкими. Процедура вычисления коэффициента полезной информации в изображении описана в Алгоритме 3. Полученный коэффициент позволяет судить о наличии шума в изображении.

**Алгоритм 3:** Вычисление коэффициента полезной информации в изображении

```
Вход: оригинал изображения bitA; модифицированное изображение bitB
Результат: коэффициент изменения изображения
information ← 0;
bitA ← new BitArray(argA);
bitB ←  new BitArray(argB);
information ← Math.Log(bitA, 2) - Math.Log(bitB, 2);
return information;
```

Качество информации в изображении связано не только с занимаемым объемом изображения, но и с вероятностью события появления пикселя того или иного цвета. Другими словами, чем больше информации содержится в изображении, тем меньше неопределенность события появления пикселя определённого цвета. Используя определение энтропии по Шеннону, вычислим качество информации, переданной в изображении при слиянии данных. Процедура вычисления коэффициента энтропии описана в Алгоритме 4. С помощью формулы (5), использованной в процедуре, определяем влияние наличия данных другого изображения на его качество.

**Алгоритм 4:** Вычисление энтропии по Шеннону

```
Вход: изображение imgVolume
Результат: коэффициент энтропии
result ← 0;
total ← 0;
freq ← new List<byte, int>();
for i ∈ imgVolume.Length do
  for j ∈ imgVolume.Length do
        color ← imgVolume[i, j]
        freq[color] ← freq[color] + 1
        total ← total + 1
  end
end
for b ∈ freq.Length do
  if freq[b] != 0 then
        prob ← freq[b] / total
        result ← result – prob * Math.log(prob, 2);
  end
end
return result;
```

На основе формулы (6) процедура принятия решения описана в Алгоритме 5. Процедура принимает на вход данные о качестве изображения, полученные от процедур вычисления глубины изображения, коэффициентов полезной информации в изображении и коэффициентов энтропии. На выходе будут получены результаты по каждой формуле оценки качества изображения.

**Алгоритм 5:** Вычисление взвешенного решения о качестве изображения

```
Вход: весовые коэффициенты double Wn, Wi, Wh; пороговые
значения double Tmax, Tmin; количество цветов в палитре
double N; коэффициента полезной информации в изображении
double I; коэффициент энтропии double H;
Результат: результат оценки качества изображения
Q ← -1;
result ← (Wn * N) + (Wi * I) + (Wh * H)
if result >= Tmax then
  Q ← 1;
else
  if result >= Tmin AND result < Tmax then
        Q ← 0;
  end
end
return Q;
```

**Результаты и обсуждение.** Для расчета информативности и оценки качества изображений применим перечисленные выше алгоритмы. В качестве исходных данных использовались изображения из набора данных RoadScene, содержащего инфракрасные и видимые изображения сцен дорог, транспортных средств и пешеходов. Изображения имеют одинаковый размер и одинаковый ракурс. Данные находятся в открытом доступе

по адресу https://github.com/hanna-xu/RoadScene. На рис. 2 представлены два оригинальных снимка полученных с разных приборов, снимок в видимом диапазоне (рис. 2,а) и в инфракрасном диапазоне (рис. 2,б).

а б

*Рис. 2. Снимки участка дороги в двух световых диапазонах*

В результате объединения происходит слияние информативных частей в единое изображение. Для того чтобы оценить, насколько новое изображение стало более информативным, насколько адекватны данные, уместны или применимы данные, привносимые вторым изображением, вычислим энтропию. Преобладающая часть методов слияния данных основывается на предположении о том, что входные данные имеют одинаковую достоверность и играют симметричную роль. Однако, входные данные могут иметь разную достоверность и вносить разную информацию в оригинальное изображение. Для того, чтобы не было снижения качества результатов слияния, необходимо учитывать данный факт, поэтому следует вычислять достоверность информации на всех этапах слияния. Формула (6) позволяет оценить все возможные варианты и принять решение на наилучшем варианте. Для начала приведем все изображения к градациям серого с глубиной цвета 8 бит. На примере данных выполним расчеты информативности изображений в видимом и инфракрасном диапазонах.

а б в

*Рис. 3. Изображения участка дороги в градациях серого*

На рис. 3 приведены изображения в видимом (рис. 3,а) и инфракрасном диапазоне (рис. 3,б) и результат слияния данных изображений (рис. 3,в). Все изображения имеют разрешение 500 × 329 пикселей. Были вычислены энтропии каждого изображения согласно формуле (4), так же были вычислены значения энтропии по изображениям, полученным при слиянии в видимом и инфракрасном диапазоне. Значения полученных результатов вычислений приведены в табл, 1.

Таблица 1

**Результаты вычисления коэффициентов энтропии**

| | Видимый диапазон | Инфракрасный диапазон | Слияние видимого и инфракрасного диапазонов |
|---|---|---|---|
| Коэффициент энтропии по Шеннону | 6.499 | 6.962 | 7,093 |

Из табл. 1 можно видеть, что наилучший результат получен при слиянии изображений в видимом и инфракрасном диапазонах. Вычислим для каждого из изображений коэффициент качества изображения согласно формуле (6). Для параметров $N$, $I$, $H$ используем весовые коэффициенты $w_N = 0.4$, $w_I = 0.3$ и $w_H = 0.3$. Определим пороговые значения $T_{\max}$ и $T_{\min}$ равным 2.6 и 1.8, соответственно. Результаты вычислений представлены в табл. 2.

Таблица 2

**Результаты вычисления мер качества изображений**

| | Видимый диапазон | Инфракрасный диапазон | Слияние видимого и инфракрасного диапазонов |
|---|---|---|---|
| Уровень качества изображения | 2.509 | 2.638 | 2.691 |
| Q | 0 | 1 | 1 |

Рассмотрим надежность подхода в случае, когда одно из изображений имеет размытие, и выходной результат имеет шум. Было взято изображение, полученное при слиянии видимого и инфракрасного диапазона, и получены варианты изображений с добавлением размытия по Гауссу в 2 пикселя, 4 пикселя, 6 пикселей и 8 пикселей. Так же варианты с добавлением шума, по Гауссу – 10 %, 20 %, 30 % и 40 %. Результат вычисления коэффициентов энтропии приведен в табл. 3.

Таблица 3

**Результаты влияния шумов и размытости на энтропию объединённого изображения**

| | Размытие 0 пикселей | Размытие 2 пикселя | Размытие 4 пикселя | Размытие 6 пикселей | Размытие 8 пикселей |
|---|---|---|---|---|---|
| Шум 0% | 7.093 | 7.068 | 7.040 | 7.015 | 6.997 |
| Шум 10% | 7.421 | 7.410 | 7.396 | 7.381 | 7.373 |
| Шум 20% | 7.667 | 7.660 | 7.648 | 7.636 | 7.633 |
| Шум 30% | 7.801 | 7.797 | 7.794 | 7.792 | 7.790 |
| Шум 40% | 7.831 | 7.833 | 7.837 | 7.839 | 7.840 |

Из полученных результатов в табл. 3 видно, что значение энтропии по Шеннону уменьшается с увеличением размытости и увеличением шумов, что свидетельствует об уменьшении качества информативности изображения. Дополнительное применение коэффициента полезной информации в изображении по Хартли также позволяет судить о количестве шумов в изображении для получения оценки качества изображений.

В случае анализа цветного изображения подсчет информативности изображений следует проводить по каждому каналу – красному, зеленому и синему отдельно. Уменьшение энтропии не всегда эквивалентно увеличению информативности, как видно из табл. 3. Сглаживание приводит к уменьшению коэффициента энтропии, а увеличение шума приводит к его увеличению. Если в изображении увеличивается количество шума, то сглаживание не оказывает эффекта уменьшения коэффициента энтропии.

**Заключение.** Использование методов слияния данных позволяет получить более информативные и содержательные результаты, повышая точность, полноту и надёжность данных для последующего анализа и использования в алгоритмах машинного обучения. Предложенный подход позволяет вычислить наилучший вариант слияния изображений, при котором результирующее изображение будет иметь наибольшую информативность. Входные данные могут иметь не одинаковую достоверность и привносить различную информативность при слиянии данных, что может привести к проблемам несогласованности или неточности данных. Вычисление энтропии по Шеннону позволяет определить наиболее подходящий вариант слияния изображений для дальнейшей обработки, не прибегая к вычислению достоверности информации, а коэффициент полезной информации по Хартли позволяет судить о количестве присутствующих шумов в изображении. Предложенный подход может быть применен при анализе изображений, получаемых путём слияния нескольких исходных изображений из различных, в том числе гетерогенных источников.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК


1. *Yadav S.P., Yadav S.,* Image fusion using hybrid methods in multimodality medical images // Med. Biol. Eng. Comput. – 2020. – Vol. 58, No. 4. – P. 669-687.
2. *Meyer-Baese A., Schmid V.* The Wavelet Transform in Medical Imaging // Pattern Recognition and Signal Analysis in Medical Imaging. – 2014. – P. 113-134.
3. *Bijelic M.* Seeing Through Fog Without Seeing Fog: Deep Multimodal Sensor Fusion in Unseen Adverse Weather // The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). – 2019. – 11 p.
4. *Xu H., Ma J., Le Z., Jiang J., & Guo X*. FusionDN: A Unified Densely Connected Network for Image Fusion // Proceedings of the AAAI Conference on Artificial Intelligence. – 2020. – Vol. 34, No. 07. – P. 12484-12491.
5. *Chernov A.V., Savvas I.K., Alexandrov A.A., Kartashov O.O., Polyanichenko D.S., Butakova M.A., Soldatov A.V.* Integrated Video and Acoustic Emission Data Fusion for Intelligent Decision Making in Material Surface Inspection System // Sensors. – 2022. – Vol. 22, No. 21. – 8554 p.
6. *Li H.* A Multi-Sensor Fusion Framework Based on Coupled Residual Convolutional Neural Networks // Remote Sens. – 2020. – Vol. 12, No. 12. – 2067 p.
7. *Rogova L., Bosse E.* Information quality in information fusion // in 2010 13th International Conference on Information Fusion, Edinburgh: IEEE. – 2010. – P. 1-8.
8. *Castanedo F.* A Review of Data Fusion Techniques // Sci. World J. – 2013. – P. 1-19.
9. *Kenda K., Kažič B., Novak E., Mladenić D* Streaming Data Fusion for the Internet of Things // Sensors. – 2019. – Vol. 19, No. 8. – 1955 p.
10. *Proppe C., Kaupp J.* On information fusion for reliability estimation with multifidelity models // Probabilistic Engineering Mechanics. – 2022. – Vol. 69. – 103291 p.
11. *Umme S., Morium A., Mohammad S.U.,* Image Quality Assessment through FSIM, SSIM, MSE and PSNR—A Comparative Study // Journal of Computer and Communications. – 2019. – Vol. 07. – P. 8-18.
12. *Kavitha S., Thyagharajan K.K.* A Survey on Quantitative Metrics for Assessing the Quality of Fused Medical Images // Res. J. Appl. Sci. Eng. Technol. – 2016. – Vol. 12, No. 3. – P. 282-293.
13. *Hao Q., Zhao Q., Sbert M., Feng Q., Ancuti C., Feixas M., Vila M*. Information-Theoretic Channel for Multi-exposure Image Fusion // The Computer Journal. – 2023. – Vol. 66. – P. 114-127.
14. *Li B., Li R., Liu Z., Li C., Wang Z.* An Objective Non-Reference Metric Based on Arimoto Entropy for Assessing the Quality of Fused Images // Entropy. – 2019. – Vol. 21, No. 9. – 879 p.
15. *Vila M., Bardera A., Feixas M., Bekaert P., Sbert M.* Analysis of image informativeness measures // 2014 IEEE International Conference on Image Processing (ICIP). – 2014. – P. 1086-1090.
16. *Коваленко П.П.* Методика информационной оценки восприятия изображений // Научно-технический вестник информационных технологий, механики и оптики. – 2008. – № 48. – С. 45-49.
17. *Обухов А.Д., Николюкин М.С.* Метод повышения информационной ценности видеоданных на основе фильтрации кадров и оценки энтропии // Научно-технический вестник информационных технологий, механики и оптики. – 2023. – 23 (3). – С. 493-499.
18. *Кузнецов Л.А., Бугаков Д.А.* Разработка меры оценки информационного расстояния между графическими объектами // Информационно-управляющие системы. – 2013. – № 1 (62). – С. 74-79.
19. *Gao P., Li Z., Zhang H.* Thermodynamics-Based Evaluation of Various Improved Shannon Entropies for Configurational Information of Gray-Level Images // Entropy. – 2018. – Vol. 20, No. 1. – 19 p.

20. *Ayunts H., Grigoryan A., Agaian S.* Novel Entropy for Enhanced Thermal Imaging and Uncertainty Quantification // Entropy. – 2024. – Vol. 26, No. 5. – 374 p.
21. *Cui Y.* No-Reference Image Quality Assessment Based on Dual-Domain Feature Fusion // Entropy. – 2020. – Vol. 22, No. 3. – 344 p.

REFERENCES

1. *Yadav S.P., Yadav S.,* Image fusion using hybrid methods in multimodality medical images, *Med. Biol. Eng. Comput.,* 2020, Vol. 58, No. 4, pp. 669-687.
2. *Meyer-Baese A., Schmid V.* The Wavelet Transform in Medical Imaging, *Pattern Recognition and Signal Analysis in Medical Imaging,* 2014, pp. 113-134.
3. *Bijelic M.* Seeing Through Fog Without Seeing Fog: Deep Multimodal Sensor Fusion in Unseen Adverse Weather, *The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR),* 2019, 11 p.
4. *Xu H., Ma J., Le Z., Jiang J., & Guo X*. FusionDN: A Unified Densely Connected Network for Image Fusion, *Proceedings of the AAAI Conference on Artificial Intelligence,* 2020, Vol. 34, No. 07, pp. 12484-12491.
5. *Chernov A.V., Savvas I.K., Alexandrov A.A., Kartashov O.O., Polyanichenko D.S., Butakova M.A., Soldatov A.V.* Integrated Video and Acoustic Emission Data Fusion for Intelligent Decision Making in Material Surface Inspection System, *Sensors,* 2022, Vol. 22, No. 21, 8554 p.
6. *Li H.* A Multi-Sensor Fusion Framework Based on Coupled Residual Convolutional Neural Networks, *Remote Sens*., 2020, Vol. 12, No. 12, 2067 p.
7. *Rogova L., Bosse E.* Information quality in information fusion, in 2010 13th International Conference on Information Fusion, Edinburgh: IEEE, 2010, pp. 1-8.
8. *Castanedo F.* A Review of Data Fusion Techniques, *Sci. World J*., 2013, pp. 1-19.
9. *Kenda K., Kažič B., Novak E., Mladenić D* Streaming Data Fusion for the Internet of Things, *Sensors,* 2019, Vol. 19, No. 8, 1955 p.
10. *Proppe C., Kaupp J.* On information fusion for reliability estimation with multifidelity models, *Probabilistic Engineering Mechanics,* 2022, Vol. 69, 103291 p.
11. *Umme S., Morium A., Mohammad S.U.,* Image Quality Assessment through FSIM, SSIM, MSE and PSNR—A Comparative Study, *Journal of Computer and Communications,* 2019, Vol. 07, pp. 8-18.
12. *Kavitha S., Thyagharajan K.K.* A Survey on Quantitative Metrics for Assessing the Quality of Fused Medical Images, *Res. J. Appl. Sci. Eng. Technol*., 2016, Vol. 12, No. 3, pp. 282-293.
13. *Hao Q., Zhao Q., Sbert M., Feng Q., Ancuti C., Feixas M., Vila M*. Information-Theoretic Channel for Multi-exposure Image Fusion, *The Computer Journal,* 2023, Vol. 66, pp. 114-127.
14. *Li B., Li R., Liu Z., Li C., Wang Z.* An Objective Non-Reference Metric Based on Arimoto Entropy for Assessing the Quality of Fused Images, *Entropy,* 2019, Vol. 21, No. 9, 879 p.
15. *Vila M., Bardera A., Feixas M., Bekaert P., Sbert M.* Analysis of image informativeness measures, *2014 IEEE International Conference on Image Processing (ICIP),* 2014, pp. 1086-1090.
16. *Kovalenko P.P.* Metodika informatsionnoy otsenki vospriyatiya izobrazheniy [Methodology of information assessment of image perception], *Nauchno-tekhnicheskiy vestnik informatsionnykh tekhnologiy, mekhaniki i optiki* [Scientific and technical bulletin of information technologies, mechanics and optics], 2008, No. 48, pp. 45-49.
17. *Obukhov A.D., Nikolyukin M.S.* Metod povysheniya informatsionnoy tsennosti videodannykh na osnove fil'tratsii kadrov i otsenki entropii [Method of increasing the information value of video data based on frame filtering and entropy assessment], *Nauchno-tekhnicheskiy vestnik informatsionnykh tekhnologiy, mekhaniki i optiki* [Scientific and technical bulletin of information technologies, mechanics and optics],. – 2023. – 23 (3). – S. 493-499.
18. *Kuznetsov L.A., Bugakov D.A.* Razrabotka mery otsenki informatsionnogo rasstoyaniya mezhdu graficheskimi ob"ektami [Development of a measure for assessing the information distance between graphic objects], *Informatsionno-upravlyayushchie sistemy* [Information control systems], 2013, No. 1 (62), pp. 74-79.
19. *Gao P., Li Z., Zhang H.* Thermodynamics-Based Evaluation of Various Improved Shannon Entropies for Configurational Information of Gray-Level Images, *Entropy,* 2018, Vol. 20, No. 1, 19 p.
20. *Ayunts H., Grigoryan A., Agaian S.* Novel Entropy for Enhanced Thermal Imaging and Uncertainty Quantification, *Entropy,* 2024, Vol. 26, No. 5, 374 p.
21. *Cui Y.* No-Reference Image Quality Assessment Based on Dual-Domain Feature Fusion, *Entropy,* 2020, Vol. 22, No. 3, 344 p.

Статью рекомендовал к опубликованию д.т.н., профессор В.В. Курейчик.

**Александров Александр Алексеевич** – Южный федеральный университет; e-mail: alea@sfedu.ru; г. Ростов-на-Дону, Россия; Международный исследовательский институт интеллектуальных материалов Южного федерального университета; инженер.

**Мизюков Григорий Сергеевич** – Ростовский государственный университет путей сообщения; e-mail: g.miziukov@yandex.ru; г. Ростов-на-Дону, Россия; кафедра вычислительной техники и автоматизированных систем управления; к.т.н.; доцент.

**Бутакова Мария Александровна** – Научно-исследовательский и проектно-конструкторский институт информатизации, автоматизации и связи на железнодорожном транспорте АО «НИИАС», Ростовский филиал; e-mail: m.butakova@vniias.ru; г. Ростов-на-Дону, Россия; д.т.н.; профессор; г.н.с.

**Aleksandrov Aleksandr Alekseevich** – Southern Federal University; e-mail: alea@sfedu.ru; Rostov-on-Don, Russia; the International Research Institute of Intelligent Materials of the Southern Federal University; engineer.

**Miziukov Grigorii Sergeevich** – Rostov State Transport University; e-mail: g.miziukov@yandex.ru; Rostov-on-Don, Russia; the Department of Computer Engineering and Automated Control Systems; cand. of tech. sc.; associate professor.

**Butakova Maria Aleksandrovna** – JSC NIIAS; e-mail: m.butakova@vniias.ru; dr. of tech. sc.; professor; leading researcher.

УДК 65:614.842: 005.591.1 DOI 10.18522/2311-3103-2024-5-131-142

**О.С. Малютин, Р.Ш. Хабибулин**

# МЕТОДИКА ОПРЕДЕЛЕНИЯ ЧАСТОТЫ ВОЗНИКНОВЕНИЯ ПОЖАРОВ В ЗДАНИЯХ НА ОСНОВЕ МЕТОДОВ ОЦЕНКИ ПЛОТНОСТИ И ИМИТАЦИИ ОТЖИГА

*Решение задачи определения оптимального пространственного размещения пожарных подразделений представляет собой достаточно сложную научно-техническую проблему, включающую, как показали предыдущие исследования, обширный перечень факторов, в том числе необходимость оценки ожидаемых частот возникновения пожаров в разных частях населенных пунктов в зависимости от характера застройки. В настоящее время в Российской Федерации подходы и методы, позволяющие решать эту проблему, не достаточно проработаны. Как правило исследователи ограничиваются фактом существования пространственного распределения пожаров, не углубляясь в причины, приведшие к тому или иному характеру такого распределения. Между тем их понимание позволит строить модели оценки ожидаемых плотностей потоков пожаров в различных районах городов. В статье предложен подход, основанный на методе оценки пространственной плотности случайных событий (KDE, Kernel Density Estimation) и алгоритме имитации отжига для подбора значений расчетных частот возникновения пожаров в зданиях различных классов функциональной пожарной опасности. Существующая классификация расширена за счет добавления класса Ф1.5 для садовых домиков и дач. Подход апробирован на имеющихся данных о пожарах за период 2010-2020 годов и городской застройке города Красноярск. Подход реализован в виде программного решения на языке программирования Python с использованием инструментов ГИС, пространственного и сетевого анализа. Исследование показало, что предложенный подход позволяет получить такие значения частот возникновения пожаров, при которых их прогнозируемая плотность будет максимально близка к фактической. Полученные результаты расширяют набор исследовательских инструментов в области оценки как фактической, так и прогнозируемой пожарной обстановки и направлены на развитие методов и алгоритмов определения оптимальных мест размещения пожарных подразделений. Предложенный подход также может быть использован и при решении иных задач пространственной оптимизации в области обеспечения общественной безопасности, безопасности дорожного движения, защиты населения от чрезвычайных ситуаций, а также в области урбанистики и градостроительства.*

*Пожарная охрана; частота пожара; геоинформационные системы; оптимизация; метод оценки плотности.*

**O.S. Malyutin , R.Sh. Khabibulin**

## BUILDINGS FIRES FREQUENCY DETERMINING METHODOLOGY BASED ON DENSITY ESTIMATION AND SIMULATED ANNEALING METHODS

*Solving the problem of determining the optimal spatial location of fire departments is a rather complex scientific and technical problem, including, as previous studies have shown, an extensive list of factors, including the need to assess the expected frequencies of fires in different parts of settlements, depending on the nature of the development. Currently, in the Russia, approaches and methods to solve this problem are not sufficiently developed. As a rule, researchers limit themselves to the fact of the existence of a spatial distribution of fires, without delving into the causes that led to one or another nature of such a distribution. Meanwhile, their understanding will allow us to build models for estimating the expected densities of fire flows in various areas. The article proposes an approach based on the method of estimating the spatial density of random events (KDE, Kernel Density Estimation) and a simulated annealing algorithm to select the values of the calculated frequencies of fires in buildings of various classes of functional fire hazard. The approach has been tested on the available data on fires for the period 2010-2020 and urban development in the city of Krasnoyarsk. The study showed that the proposed approach allows us to obtain such values of fire occurrence frequencies at which their predicted density will be as close as possible to the actual one. The results obtained expand the set of research tools in the field of assessing both the actual and predicted fire situation and are aimed at developing methods and algorithms for determining the optimal locations of fire departments. The proposed approach can also be used to solve other problems of spatial optimization in the field of public safety, road safety, protection of the population from emergency situations, as well as in the field of urbanism and urban planning.*

*Fire protection; fire frequency; geoinformation systems; optimization; density estimation method.*

**Введение.** При решении задачи поиска оптимального размещения пожарных подразделений, одним из ключевых факторов, который следует учитывать, является частота возникновения пожаров в тех или иных частях населенного пункта. Исследованию данной проблематики посвящены работы [1–7], в которых отражены результаты исследований связанных с оценкой частот возникновения пожаров в зданиях и на территории населенных пунктов. Основным недостатком названых работ можно назвать выборочную оценку частот возникновения пожаров лишь для ряда случаев (жилые дома, промышленные предприятия и т.д.). В статистических сборниках «Пожары и пожарная безопасность» [8, 9] приводятся статистические сведения о количестве пожаров в зданиях различного назначения и классов функциональной пожарной опасности, однако, не располагая сведениями об общем количестве зданий соответствующих классов, оценить частоты возникновения пожаров в них невозможно. Сведения о статистических данных о частотах возникновения пожара в зданиях приводятся в приложении №3 к методике определения расчетных величин пожарного риска в зданиях, сооружениях и пожарных отсеках различных классов функциональной пожарной опасности, утвержденной приказом МЧС России от 14.11.2022 № 1140 [10]. Из числа зарубежных работ стоит отметить [11], где для оценки плотности распределения пожаров используется метод оценки ядерной плотности (далее – *KDE* от англ. *Kernel Density Estimation*), а также работу [12], в которой описывается применение индекса Морана. В компьютерных имитационных системах, например [1], имеется возможность указывать плотности возникновения пожаров, но их значения выбираются самим пользователем. При этом, способов оценки ожидаемой плотности пожаров на территориях населенных пунктов, для которых сведения о пожарах недостаточны, или отсутствуют (например, при проектировании новых районов застройки), на сегодняшний день не существует.

Одним из подходов к оценке ожидаемой частоты возникновения пожаров может стать использование заранее определенных справочных значений о расчетных частотах возникновения пожаров в зданиях различного назначения.

Таким образом, работа посвящена оценке и подбору методов и справочных данных о частотах возникновения пожаров в зданиях различного назначения с точки зрения их предсказательной силы в отношении ожидаемой пространственной плотности пожаров с использованием методов пространственной аналитики и машинного обучения.

Объектом исследования являются частоты возникновения пожаров (на примере г. Красноярск). Предмет исследования – методика определения частот возникновения пожаров в зданиях различного назначения.

**Методы**

***Метод KDE.*** Метод *KDE* позволяет оценить плотность пространственного распределения некоторых случайных событий (применительно к данному исследованию – мест возникновения пожаров и мест размещения зданий). Метод был предложен Е. Парценом (в 1956 году)[13] и М. Розенблаттом (в 1962 году) [14] независимо друг от друга, и основан на идее сглаживания количественного распределения за счет оценки влияния случайных величин в некоторой близости от них по некоторому закону. Метод может быть взвешенным или невзвешенным, что удобно как для оценки пространственной плотности мест возникновения пожаров, так и для оценки ожидаемой плотности возникновения пожаров в зданиях различного назначения. В качестве закона распределения влияния случайных величин могут использоваться следующие функции (ядра): Гаусса, tophat, Епанечникова, экспоненциальная, линейная и косинусная [15].

В данной работе использовалась программная реализация метода *KDE* из библиотеки языка программирования *Python Scikit-learn* версии 1.4.2.

***Алгоритм имитации отжига.*** Алгоритм имитации отжига (*англ. – simulated annealing*, далее – ИО) – это метод оптимизации, основанный на идее термодинамической системы, охлаждающейся до состояния равновесия. В процессе работы алгоритма случайным образом выбирается новое решение задачи, которое сравнивается с текущим. Если новое решение лучше текущего, оно принимается как текущее. Если же новое решение хуже, то оно всё равно может быть принято, как текущее с некоторой вероятностью, зависящей от температуры системы. По мере того, как температура снижается в ходе вычислений, вероятность принятия худшего решения уменьшается, и система приближается к глобальному оптимуму [16].

**Исходные данные**

***Данные о пожарах.*** Для проведения анализа частоты возникновения пожаров в качестве исходных данных были использованы сведения из Федеральной базы данных (ФБД) «Пожары» за период 2010-2020 годы. Была использована выборка по пожарам, произошедшим в зданиях на территории города Красноярск. Пожары, произошедшие на открытой территории (загорания мусора, пожары транспорта и т.д.) не рассматривались. Для анализа были выбраны следующие признаки: дата пожара, класс функциональной пожарной опасности (далее – КФПО), назначение объекта, адрес объекта. Поскольку в ФБД «Пожары» сведения о географических координатах отсутствуют, была проведена процедура геокодирования с использованием возможностей сервиса «Яндекс Геокодер». Часть пожаров произошла на территории садоводческих товариществ и дачных поселков, провести процедуру геокодирования для которых не представляется возможным в связи с отсутствием данных о точных адресах дачных построек – такие пожары были расположены случайным образом на территории садоводческих товариществ.

В результате был получен набор данных, состоящий из 5 620 записей вида:

Таблица 1

**Пример исходных данных о пожарах**

| id | Дата | Идентификатор объекта по 1140 | КФПО | Пожарных автомобилей всего | Адрес | Координаты |
|---|---|---|---|---|---|---|
| 0 | 2011-12-14 02:46:00 | 25 | Ф1.4 | 1 | ул. Большевистская 175/5. | POINT (82.973415, 54.994442) |
| 1 | 2010-11-29 11:23:00 | 40 | Ф1.4 | 2 | ул. Жуковского 106/1. | POINT (82.893686, 55.055223) |
| 2 | 2019-10-21 00:58:00 | 16 | Ф1.3 | 6 | ул. Восточная 26. | POINT (82.890907, 55.055125) |

***Данные о застройке города.*** В качестве исходных данных о застройке города были использованы данные из открытого картографического сервиса *Open street map* (далее – *OSM*). Получение данных осуществлялось посредством библиотеки *Python OSMNX* [17, 18]. Полученные данные были сохранены в формате пространственных данных *geopackage* и включены в проект приложения *QGIS* [19].

Сведения о назначении зданий для обеспечения дальнейшего сопоставления со значениями КФПО были соответствующим образом интерпретированы.

В результате было получено 85 094 записи вида:

Таблица 2

**Пример исходных данных о зданиях**

| id | Назначение | Адрес | Этажей | КФПО |
|---|---|---|---|---|
| 207766295 | Жилой дом | ул. Береговая, 12 | 1 | Ф1.4 |
| 791957821 | Административное здание | ул. Песочная, 2 | 5 | Ф4.3 |
| 1247655719 | Торговое здание | ул. Мира 22 | 3 | Ф3.1 |

**Обзор данных.** Записи о произошедших пожарах были размещены на интерактивной карте *QGIS* в виде точечных объектов в соответствии с полученными координатами. Далее встроенными средствами *QGIS* была построена тепловая карта пожаров для ячеек размером 25 м и зоной охвата 1 км (рис. 1).

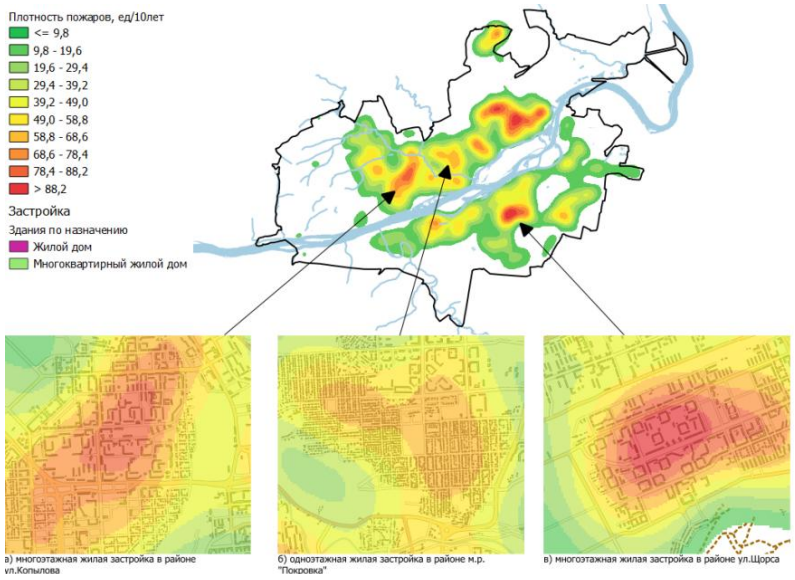



*Рис. 1. Распределение плотностей возникновения пожаров на территории города Красноярск за период 2010- 2020 гг.*

Обзор полученного распределения показывает наличие стойких паттернов пространственного распределения пожаров, связанных с характером застройки. Наибольшая плотность пожаров наблюдается в районах с многоэтажной жилой застройкой (78,4 пожара за 10 лет) (рис. 1 а,в). Несколько меньшая (58,8–78,4 пожара за 10 лет) плотность наблюдается в районах с одноэтажной жилой застройкой (рис. 3,б).

На следующем этапе была проведена оценка соответствия вероятностей возникновения пожаров, приведенная в методике определения расчетных величин пожарного риска в зданиях, сооружениях и пожарных отсеках различных классов функциональной пожарной опасности [10]. Для этого зданиям, в зависимости от их назначения, было сопоставлено значение вероятности возникновения пожара. При этом, для ряда зданий, сопос-

тавить которые с конкретным классом не представлялось возможным, было установлено среднее значение по всем классам (0,015 ед/год). На основании полученных значений была составлена тепловая карта ожидаемой плотности возникновения пожаров с размером ячейки 25 м и зоной охвата 1 км (рис. 2). Для того, чтобы плотности можно было сопоставить между собой их значения были нормализованы, т.е. приведены к общему диапазону в пределах от 0 до 1, где 0 соответствует минимальному значению плотности, а 1 – максимальному.

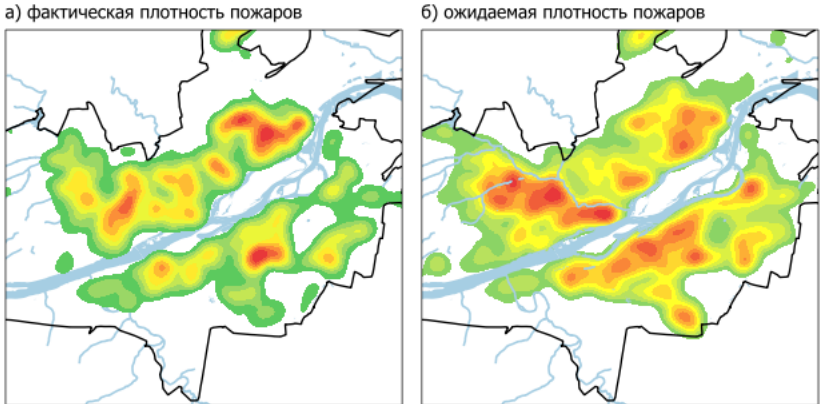



*Рис. 2. Сравнение плотностей пожаров: а – фактическая плотность пожаров, б – ожидаемая плотность пожаров для вероятностей возникновения пожаров в соответствии с методикой [10]*

Из рис. 2 видно, что фактическая и ожидаемая в соответствии со значениями риска возникновения пожара по методике [10], плотности пожаров не совпадают. Частоты возникновения пожаров в зданиях различного назначения оказывают существенное влияние на распределение плотностей пожаров, что говорит о том, что зная вероятности возникновения пожаров в зданиях различного назначения, можно предсказать какова будет плотность пожаров в тех или иных районах города и на основе этого прогноза вырабатывать управленческие решения по определению мест размещения пожарных депо.

**Постановка задачи исследования.** Задача частного исследования была поставлена следующим образом: требуется найти такую классификацию зданий $C$ и частоты возникновения пожаров в каждом из классов $p_{кл}$ (далее – Частотной модели), при которых значение некоторой функции ошибки $err$ между фактической $P_ф$ и ожидаемой $P_o$ плотностями пожаров, нормализованных к диапазону (0:1) в некотором множестве контрольных точек $I$ выбранных в пространстве $S$ территории населенного пункта будет минимальна:

$$P_ф = \{p_{ф1}, p_{фi}\}, i \in I, \quad (1)$$

$$P_o = \{p_{o1}, p_{oi}\}, i \in I, \quad (2)$$

$$I \subseteq S,$$

$$err(P_ф, P_o) \to min, \quad (3)$$

где $p_{фi}$ – фактическая плотность пожаров в точке $i$; $p_{oi}$ – ожидаемая плотность пожаров в точке $i$.

Для определения значений частоты возникновения пожаров в контрольных точках $I$ была использована модель ядерной оценки плотности $KDE$ [15]. Тогда фактическая плотность пожаров в любой $i$-й точке будет:

$$p_{фi} = KDE(i, F) \quad (4)$$

$$F \subseteq S,$$

где $F$ – точки возникновения пожаров.

А ожидаемая плотность пожаров в каждой точке:

$$p_{oi} = KDE(i, J, P_{зд}), \quad (5)$$

$$P_{зд} = \{p_{зд1}, p_{здj}\}, j \in J, \quad (6)$$

$$J \subseteq S, \quad (7)$$

$$p_{здj} = p_{кл}(c_j), c_j \in C, \quad (8)$$

где, $J$ – точки центров зданий; $P_{зд}$ – частоты возникновения пожаров; $p_{здj}$ – частота возникновения пожара в каждом из $j$-х зданий; $c_j$ – класс $j$-го здания; $C$ – множество возможных классов зданий; $p_{кл}$ – частота возникновения пожара в каждом из $C$ классов зданий.

Предлагается два подхода:

1. Выдвижение и оценка теоретически обоснованных частотных моделей – кандидатов.

2. Подбор наиболее подходящих параметров частотной модели с использованием алгоритма имитации отжига.

**Результаты исследования.** По первому подходу в качестве кандидатов были рассмотрены следующие модели (табл. 3).

Таблица 3

**Модели соотношения классификации зданий и частот возникновения пожаров**

| Обозначение | Описание | Выбор классификации и вероятностей возникновения пожаров |
|---|---|---|
| ПРВПЗ КФПО | Пространственное распределение плотности зданий, взвешенной по вычисленной вероятности возникновения пожара в зданиях различных классов функциональной пожарной опасности (ФЗ-123, ст. 32), далее – КФПО. | Классификация: по КФПО<br>Частота возникновения пожаров: $c_k = N_k^{п}/N_k^{зд}$, где $N_k^{п}$ – количество пожаров, произошедших в зданиях $k$-го КФПО, ед; $N_k^{зд}$ – количество зданий с $k$-м КФПО, ед. |
| ПРВПЗ 1140 | Пространственное распределение плотности зданий, взвешенной по вероятности возникновения пожара в зданиях различного назначения в соответствии с приказом МЧС России №1140 от 14.11.2022 | Классификация: прил. 3 к приказу МЧС России №1140 от 14.11.2022<br>Частота возникновения пожаров: прил. 3 к приказу МЧС России №1140 от 14.11.2022 |
| ПРВПЗ 1140+ | Пространственное распределение аналогично ПРВПЗ 1140, но дачные домики вынесены в отдельную категорию. | Классификация: прил. 3 к приказу МЧС России №1140 от 14.11.2022<br>Частота возникновения пожаров: аналогично ПРВПЗ 1140, для дачных домиков установлена вероятность вдвое меньше, чем для одноэтажных жилых домов. |
| ПРВФПЗ КФПО+ | Пространственное распределение вычислено среднее значение фактической плотностей пожаров в центроидах зданий каждого их классов, но дачные домики вынесены в дополнительную категорию Ф1.5. | Классификация: КФПО + Ф1.5<br>Частота возникновения пожаров: определена по реальным плотностям пожаров. |
| ПРВПЗ КФПО+1140 | Пространственное распределение плотности зданий, взвешенной по вероятности возникновения пожара в зданиях различного КФПО в соответствии с приказом 1140 плотности зданий с уточнениями | Классификация: по КФПО + Ф1.5<br>Частота возникновения пожаров: аналогично ПРВПЗ 1140, для Ф1.5, установлена вероятность вдвое меньше, чем для Ф1.4. |

Для оценки точности перечисленных моделей были применены следующие функции ошибки: MAE – средняя абсолютная ошибка (англ. – Mean Absolute Error), MSE – средняя квадратичная ошибка (англ. – Mean Square Error), R2 – коэффициент детерминации [20]. Результаты расчета приведены в табл. 4. Вычисления проводились для ячеек размером 200х200 м, охватом 1000 м и гауссовой функции оценки плотности.

Таблица 4

**Результаты вычисления функций ошибки для моделей соотношения классификации зданий и частот возникновения пожаров**

| Обозначение | Функция ошибки | | |
|---|---|---|---|
| | MAE | MSE | R2 |
| ПРВПЗ КФПО | 0,065929 | 0,012883 | 0,791836 |
| ПРВПЗ 1140 | 0,06926 | 0,009375 | 0,857428 |
| ПРВПЗ 1140+ | 0,06458 | 0,008229 | 0,874893 |
| ПРВФПЗ КФПО+ | 0,10839 | 0,02803 | 0,153392 |
| ПРВПЗ КФПО+1140 | 0,077557 | 0,011606 | 0,836134 |

Графическое представление результатов расчета представлено на рис. 3.

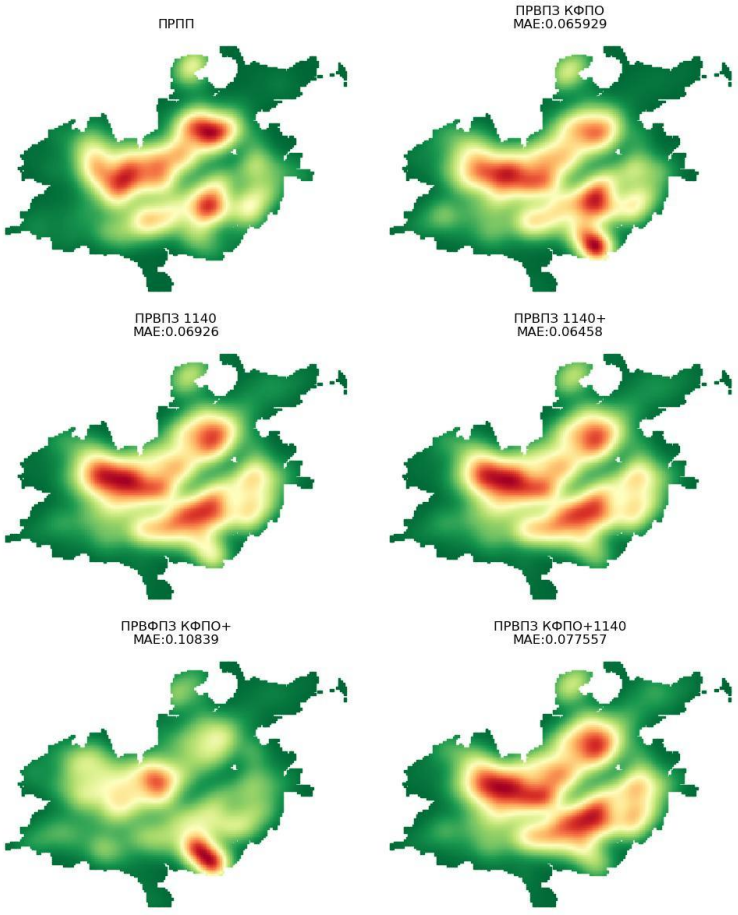



*Рис. 3. Сравнение пространственного распределения фактической (ПРПП) и ожидаемой плотностей возникновения пожаров для моделей ПРВПЗ КФПО, ПРВПЗ 1140, ПРВПЗ 1140+, ПРВФПЗ КФПО+, ПРВПЗ КФПО+1140*

Видно, что наиболее точной является ПРВПЗ 1140+ (MAE: 0,06458, MSE: 0,008229, R2: 0,874893), данная частотная модель послужила основой для дальнейшего уточнения.

Обозначение лучше поставить Вариант 1, Вариант 2 и т.д. А не ПРВПЗ КФПО!

**Поиск с использованием алгоритма имитации отжига.** Дальнейшее уточнение вероятностей возникновения пожаров в зданиях различного назначения проводилось с использованием алгоритма ИО.

Применительно к поставленной задаче охлаждающейся системой является частотная модель КФПО+, состоянием системы являются численные значения вероятностей возникновения пожаров в зданиях различных КФПО+, энергией системы является значение, возвращаемое функцией 1.4.

Гиперпараметрами алгоритма ИО являются:

- размер ячейки: *cell;*
- охват: *band_width;*
- функции оценки плотности: *kernel;*
- температура: *T.*

Алгоритм реализован с использованием языка программирования *Python*.

В табл. 5 приведены оценки точности и качества некоторых рассмотренных в работе реализаций ИО в зависимости от гиперпараметров модели. На рис. 4 они показаны графически.

Таблица 5

**Оценка точности некоторых рассмотренных настроек алгоритма ИО**

| Метрика | ***ИО М0:*** *cell=200м, band_width = 1000м, kernel = 'gaussian', T = 0,0001* | ***ИО М1:*** *cell=200м, band_width = 400м, kernel = 'tophat', T = 0,0001* | ***ИО М2:*** *cell=200м, band_width = 1000м, kernel = 'tophat', T = 0,0001* |
|---|---|---|---|
| MAE | 0,071375 | 0,056051 | 0,057503 |
| MSE | 0,008966 | 0,00607 | 0,006656 |
| R2 | 0,865552 | 0,909139 | 0,886433 |
| Время расчета, час | 32 | 3 | 3 |

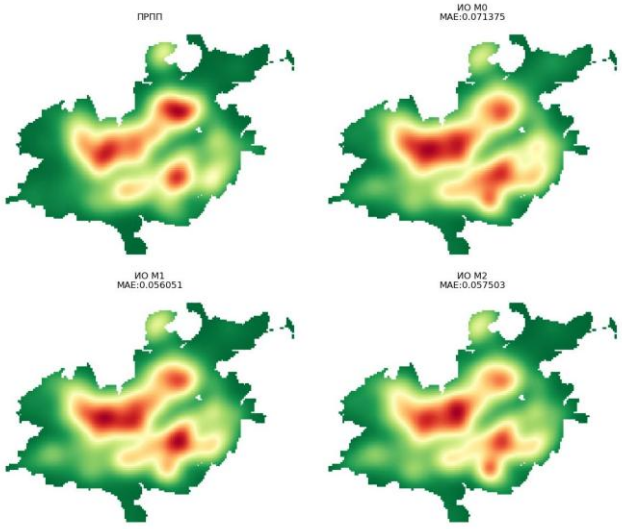



*Рис. 4. Сравнение пространственного распределения фактической (ПРПП) и ожидаемой плотности возникновения пожаров для моделей ИО М0, ИО М1, ИО М2*

В табл. 6 приведены результаты моделирования вероятностей возникновения пожаров в зданиях различного КФПО. Полученные результаты сопоставлены с фактическими значениями.

Таблица 6

**Сравнение вычисленных и фактических значений частоты возникновения пожаров в зданиях различных КФПО**

| КФПО | Количество зданий, ед. | Вычисленная частота, пож/год | Фактическая частота, пож/год |
|---|---|---|---|
| Ф1.1 | 605 | $5{,}12 \cdot 10^{-3}$ | $9 \cdot 10^{-4}$ |
| Ф1.2 | 565 | $6{,}01 \cdot 10^{-3}$ | $5{,}72 \cdot 10^{-3}$ |
| Ф1.3 | 6059 | $1{,}68 \cdot 10^{-2}$ | $4{,}7 \cdot 10^{-2}$ |
| Ф1.4 | 18587 | $2 \cdot 10^{-3}$ | $5{,}2 \cdot 10^{-3}$ |
| Ф1.5 | 13463 | $5{,}99 \cdot 10^{-5}$ | $2 \cdot 10^{-3}$ |
| Ф2.1 | 35 | $1{,}64 \cdot 10^{-3}$ | $2 \cdot 10^{-2}$ |
| Ф2.2 | 7 | $3{,}9 \cdot 10^{-3}$ | $5{,}19 \cdot 10^{-2}$ |
| Ф2.3* | 0 | $4{,}21 \cdot 10^{-3}$ | 0 |
| Ф2.4* | 0 | $4{,}21 \cdot 10^{-3}$ | 0 |
| Ф3.1 | 2902 | $7{,}24 \cdot 10^{-3}$ | $9{,}49 \cdot 10^{-3}$ |
| Ф3.2 | 173 | $7{,}4 \cdot 10^{-3}$ | $2{,}68 \cdot 10^{-2}$ |
| Ф3.3 | 14 | $6{,}7 \cdot 10^{-3}$ | 0 |
| Ф3.4* | 0 | $4{,}21 \cdot 10^{-3}$ | 0 |
| Ф3.5* | 0 | $4{,}21 \cdot 10^{-3}$ | 0 |
| Ф3.6 | 87 | $1 \cdot 10^{-2}$ | $1{,}9 \cdot 10^{-2}$ |
| Ф3.7 | 61 | $6{,}35 \cdot 10^{-3}$ | $2{,}98 \cdot 10^{-3}$ |
| Ф4.1 | 222 | $6 \cdot 10^{-3}$ | $4{,}5 \cdot 10^{-3}$ |
| Ф4.2 | 180 | $6{,}4 \cdot 10^{-3}$ | $5 \cdot 10^{-4}$ |
| Ф4.3 | 5794 | $1{,}54 \cdot 10^{-3}$ | $1{,}22 \cdot 10^{-3}$ |
| Ф4.4 | 11 | $7{,}42 \cdot 10^{-3}$ | 0 |
| Ф5.1 | 3767 | $1{,}42 \cdot 10^{-3}$ | $5{,}79 \cdot 10^{-3}$ |
| Ф5.2 | 4035 | $4{,}59 \cdot 10^{-3}$ | $1{,}15 \cdot 10^{-2}$ |
| Ф5.3* | 0 | $4{,}21 \cdot 10^{-3}$ | 0 |
| Ф0 | 1525 | $3{,}81 \cdot 10^{-4}$ | $1{,}26 \cdot 10^{-2}$ |

** примечание – для ряда КФПО теоретическую частоту возникновения пожаров установить не удалось в связи с отсутствием сведений об объектах соответствующего класса на территории города Красноярска. Для таких зданий установлено среднее значение частоты возникновения пожаров для известных КФПО.*

Из табл. 6 видно, что наибольшая вероятность возникновения пожаров – в зданиях класса Ф1.3 (многоквартирные жилые дома), наименьшая – в зданиях Ф1.5 (дачные домики, предназначенные для сезонного проживания).

**Обсуждение.** Необходимо отметить, что полученные теоретические и фактические значения для ряда классов объектов защиты заметно отличаются, что возможно стало следствием малого количества пожаров в зданиях соответствующих классов. Это не по-

зволяет на данном этапе построить достаточно точную модель определения вероятности возникновения пожаров. Однако, как видно из табл. 4 и рис. 3, модели ПРВПЗ КФПО и ПРВПЗ КФПО+ построенные на вычисленных значениях вероятностей возникновения пожаров уступают по точности построения пространственных распределений ожидаемой плотности пожаров моделям полученным в результате применения алгоритма имитации отжига (ИО М0, ИО М1, ИО М2) (табл. 5, рис. 4).

Существенное влияние на результаты работы может оказывать качество исходных данных – данные о пожарах и данные о городской застройке. Использованные в работе данные *OSM* не содержат в полной мере всех необходимых для анализа данных. Так, например, использование в сервисе собственной системы идентификации объектов, позволяющей к тому же указывать любые значения помимо принятых, приводит к необходимости интерпретации имеющихся данных с точки зрения принятой в России пожарно-технической классификации. Заметная доля объектов не имеет или имеет ошибочные сведения о назначении объектов, этажности, адресах и т.д. Все это приводит к необходимости трудоемкой ручной обработки данных, что делает невозможным использование данных *OSM* в чистом виде.

**Заключение.** В ходе работы предложен и апробирован метод оценки частоты возникновения пожаров в зданиях различного назначения на основе пространственной плотности пожаров и алгоритма имитации отжига. Установлено, что предложенный метод позволяет определить частоты возникновения пожаров, на основе которых можно получить максимально согласующиеся с реальностью ожидаемые плотности пожаров на территории городов. А это в свою очередь позволяет оценить оптимальность размещения существующих и необходимость создания новых пожарных подразделений на территории городов с целью повышения защиты населения от пожаров.

Направления дальнейших исследований:

- оценка перспектив замены использованного алгоритма имитации отжига на иные алгоритмы машинного обучения;
- разработка прикладной методики предварительной подготовки данных для использования с предложенным методом;
- реализация предложенного метода в виде программного обеспечения для ЭВМ;
- оценка необходимости включения в учет дополнительных параметров, таких как площадь и этажность здания, социальные факторы и т.д.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Соколов С.В., Брушлинский Н.Н., Фам К.Х.* Разработка и адаптация имитационной системы оперативной деятельности пожарных подразделений к условиям Вьетнама // Пожары и чрезвычайные ситуации: предотвращение, ликвидация. – 2021. – № 2. – С. 5-14. – DOI: 10.25257/FE.2021.2.5-14.
2. *Брушлинский Н.Н. [и др.]* Безопасность городов. Имитационное моделирование городских процессов и систем. – М.: ФАЗИС, 2004. – 172 с.
3. *Брушлинский Н.Н. [и др.].* Математические методы и модели управления в Государственной противопожарной службе: учебник. – 2-е изд., испр. и допол. – М.: Академия ГПС МЧС России, 2019. – 194 с.
4. *Гордиенко Д.М. [и др.].* Данные о частотах возникновения пожаров и пожароопасных ситуаций в общественных зданиях различного назначения и на производственных объектах // Пожарная безопасность. – 2009. – № 2. – С. 42-46.
5. *Кожевников М.Л.* Моделирование временных характеристик процесса функционирования пожарно-спасательных подразделений и анализ частоты использования пожарной техники // Пожары и чрезвычайные ситуации: предотвращение, ликвидация. – 2021. – № 2. – С. 79-86. – DOI: 10.25257/FE.2021.2.79-86.
6. *Власов К.С. [и др.].* Оценка распределения выездов пожарно-спасательных подразделений на пожары различных объектов. – Железногорск: Сибирская пожарно-спасательная академия Государственной противопожарной службы Министерства Российской Федерации по делам гражданской обороны, чрезвычайным ситуациям и ликвидации стихийных бедствий, 2021. – С. 574-581.
7. *Власов К.С. [и др.].* Применение технологий машинного обучения для исследования характеристик пожаров // Сибирский пожарно-спасательный вестник. – 2023. – 2 (29). – С. 80-87.

8. Пожары и пожарная безопасность в 2021 году: статист. сб. – Балашиха: ФГБУ ВНИИПО МЧС России, 2022. – 114 с.
9. Пожары и пожарная безопасность в 2022 году: статист. сб. – Балашиха: ФГБУ ВНИИПО МЧС России, 2023. – 80 с.
10. Об утверждении методики определения расчетных величин пожарного риска в зданиях, сооружениях и пожарных отсеках различных классов функциональной пожарной опасность: Приказ МЧС России от 14 ноября 2022 г. № 1140.
11. *Mazur Robert.* Assessment of Safety Level in the Aspect of 2000-2012 Fire Statistics. Temporal And Spatial Characteristics of Residential Buildings Fires in Geographical Information System. Warsaw Case Study // Bezpieczeństwo i Technika Pożarnicza. – 2014. – No. 34. – P. 47-56.
12. *Chen C.-Y., Yang Q.-H.* Hotspot Analysis of the Spatial and Temporal Distribution of Fires // Proceedings of the 4th International Conference on Geographical Information Systems Theory, Applications and Management. – Funchal, Madeira, Portugal: SCITEPRESS - Science and Technology Publications, 2018. – P. 15-21.
13. *Parzen E.* On Estimation of a Probability Density Function and Mode // Ann. Math. Statist. – 1962. – Vol. 33, No. 3. – P. 1065-1076.
14. *Rosenblatt M.* Remarks on Some Nonparametric Estimates of a Density Function // Ann. Math. Statist. – 1956. – Vol. 27, No. 3. – P. 832-837.
15. *Fan T. et al.* Density peaks clustering algorithm based on kernel density estimation and minimum spanning tree // IJICA. – 2022. – Vol. 13, No. 5/6. – P. 336.
16. *Kirkpatrick S., Gelatt C.D., Vecclii M.P.* Optimization by Simulated Annealing // Science, New Series. – 1983. – Vol. 220, No. 4598. – P. 671-680.
17. *Boeing G.* OSMnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks // Computers, Environment and Urban Systems. – 2017. – Vol. 65. – P. 126-139.
18. *Boeing G., Ha J.* Resilient by design: Simulating street network disruptions across every urban area in the world // Transportation Research Part A: Policy and Practice. – 2024. – Vol. 182. – P. 104-116.
19. *Матушкин А.С.* Картографирование и анализ пространственных данных с использованием геоинформационной системы QGIS: учеб. пособие. – Киров: ВятГУ, 2018. – 100 с.
20. *Матеров Е.Н., Бабенышев С.В.* Методы оптимизации: учеб. пособие. – Железногорск: Сибирская пожарно-спасательная академия ГПС МЧС России, 2019. – 135 с.

REFERENCES

1. *Sokolov S.V., Brushlinskiy N.N., Fam K.Kh.* Razrabotka i adaptatsiya imitatsionnoy sistemy operativnoy deyatel'nosti pozharnykh podrazdeleniy k usloviyam V'etnama [Development and adaptation of the simulation system of operational activities of fire departments to the conditions of Vietnam], *Pozhary i chrezvychaynye situatsii: predotvrashchenie, likvidatsiya* [Fires and emergencies: prediction, liquidation], 2021, No. 2, pp. 5-14. DOI: 10.25257/FE.2021.2.5-14.
2. *Brushlinskiy N.N. [i dr.]*. Bezopasnost' gorodov. Imitatsionnoe modelirovanie gorodskikh protsessov i system [The safety of cities. Simulation of urban processes and systems]. Moscow: FAZIS, 2004, 172 p.
3. *Brushlinskiy N.N. [i dr.].* Matematicheskie metody i modeli upravleniya v Gosudarstvennoy protivopozharnoy sluzhbe: uchebnik [Mathematical methods and management models in the State Fire Service: Textbook]. 2nd ed. Moscow: Akademiya GPS MChS Rossii, 2019, 194 p.
4. *Gordienko D.M. [i dr.].* Dannye o chastotakh vozniknoveniya pozharov i pozharoopasnykh situatsiy v obshchestvennykh zdaniyakh razlichnogo naznacheniya i na proizvodstvennykh ob"ektakh [Data on the frequency of fires and fire-hazardous situations in public buildings for various purposes and at industrial facilities], *Pozharnaya bezopasnost'* [Fire safety], 2009, No. 2, pp. 42-46.
5. *Kozhevnikov M.L.* Modelirovanie vremennykh kharakteristik protsessa funktsionirovaniya pozharno-spasatel'nykh podrazdeleniy i analiz chastoty ispol'zovaniya pozharnoy tekhniki [Modeling of the time characteristics of the functioning of fire and rescue units and analysis of the frequency of use of fire equipment], *Pozhary i chrezvychaynye situatsii: predotvrashchenie, likvidatsiya* [Fires and emergencies: prediction, liquidation], 2021, No. 2, pp. 79-86. DOI: 10.25257/FE.2021.2.79-86.
6. *Vlasov K.S. [i dr.].* Otsenka raspredeleniya vyezdov pozharno-spasatel'nykh podrazdeleniy na pozhary razlichnykh ob"ektov [Assessment of the distribution of visits of fire and rescue units to fires of various facilities]. Zheleznogorsk: Sibirskaya pozharno-spasatel'naya akademiya Gosudarstvennoy protivopozharnoy sluzhby Ministerstva Rossiyskoy Federatsii po delam grazhdanskoy oborony, chrezvychaynym situatsiyam i likvidatsii stikhiynykh bedstviy, 2021, pp. 574-581.
7. *Vlasov K.S. [i dr.].* Primenenie tekhnologiy mashinnogo obucheniya dlya issledovaniya kharakteristik pozharov [The use of machine learning technologies to study the characteristics of fires], *Sibirskiy pozharno-spasatel'nyy vestnik* [Siberian Fire and Rescue Bulletin], 2023, 2 (29), pp. 80-87.

8. Pozhary i pozharnaya bezopasnost' v 2021 godu: statist. sb. [Fires and fire safety in 2021: a statistical collection]. Balashikha: FGBU VNIIPO MChS Rossii, 2022, 114 p.
9. Pozhary i pozharnaya bezopasnost' v 2022 godu: statist. sb. [Fires and fire safety in 2022: a statistical collection]. Balashikha: FGBU VNIIPO MChS Rossii, 2023, 80 p.
10. Ob utverzhdenii metodiki opredeleniya raschetnykh velichin pozharnogo riska v zdaniyakh, sooruzheniyakh i pozharnykh otsekakh razlichnykh klassov funktsional'noy pozharnoy opasnost': Prikaz MChS Rossii ot 14 noyabrya 2022 g. № 1140 [On approval of the methodology for determining the calculated values of fire risk in buildings, structures and fire compartments of various classes of functional fire hazard: Order of the Ministry of Emergency Situations of Russia dated November 14, 2022 No. 1140].
11. *Mazur Robert.* Assessment of Safety Level in the Aspect of 2000-2012 Fire Statistics. Temporal And Spatial Characteristics of Residential Buildings Fires in Geographical Information System. Warsaw Case Study, *Bezpieczeństwo i Technika Pożarnicza,* 2014, No. 34, pp. 47-56.
12. *Chen C.-Y., Yang Q.-H.* Hotspot Analysis of the Spatial and Temporal Distribution of Fires, *Proceedings of the 4th International Conference on Geographical Information Systems Theory, Applications and Management*. Funchal, Madeira, Portugal: SCITEPRESS - Science and Technology Publications, 2018, pp. 15-21.
13. *Parzen E.* On Estimation of a Probability Density Function and Mode, *Ann. Math. Statist*., 1962, Vol. 33, No. 3, pp. 1065-1076.
14. *Rosenblatt M.* Remarks on Some Nonparametric Estimates of a Density Function, *Ann. Math. Statist*., 1956, Vol. 27, No. 3, pp. 832-837.
15. *Fan T. et al.* Density peaks clustering algorithm based on kernel density estimation and minimum spanning tree, *IJICA,* 2022, Vol. 13, No. 5/6, pp. 336.
16. *Kirkpatrick S., Gelatt C.D., Vecclii M.P.* Optimization by Simulated Annealing, *Science, New Series,* 1983, Vol. 220, No. 4598, pp. 671-680.
17. *Boeing G.* OSMnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks, *Computers, Environment and Urban Systems,* 2017, Vol. 65, pp. 126-139.
18. *Boeing G., Ha J.* Resilient by design: Simulating street network disruptions across every urban area in the world, *Transportation Research Part A: Policy and Practice,* 2024, Vol. 182, pp. 104-116.
19. *Matushkin A.S.* Kartografirovanie i analiz prostranstvennykh dannykh s ispol'zovaniem geoinformatsionnoy sistemy QGIS: ucheb. posobie [Mapping and analysis of spatial data using the geographic information system QGIS: tutorial]. Kirov: VyatGU, 2018, 100 p.
20. *Materov E.N., Babenyshev S.V.* Metody optimizatsii: ucheb. posobie [Optimization methods: a textbook]. Zheleznogorsk: Sibirskaya pozharno-spasatel'naya akademiya GPS MChS Rossii, 2019, 135 p.

Статью рекомендовал к опубликованию д.ф.-м.н., профессор Г.В. Куповых.

**Малютин Олег Сергеевич** – Сибирская пожарно-спасательная академия ГПС МЧС России; e-mail: malyutin@sibpsa.ru; г. Железногорск, Россия; тел.: +79835752398; отдел информационных технологий и компьютерного моделирования; начальник.

**Хабибулин Ренат Шамильевич** – Академия Государственной противопожарной службы МЧС России; e-mail: kh-r@yandex.ru, г. Москва, Россия; учебно-научный комплекс автоматизированных систем и информационных технологий; начальник.

**Malyutin Oleg Sergeevich** – Siberian Fire-Rescue academy; e-mail: malyutin@sibpsa.ru; Zheleznogorsk, Russia; phone: +79835752398; Informational Technology and Computer Modelling Branch; chief.

**Habibulin Renat Shamil'evich** – State Fire Academy; e-mail: kh-r@yandex.ru; Moscow, Russia; Educational and Scientific Complex of Automated Systems and Information Technologies; chief.

УДК 004.89 DOI 10.18522/2311-3103-2024-5-143-151

**Ж. Мохаммад**

# ИЗВЛЕЧЕНИЕ КЛЮЧЕВЫХ ФРАЗ НА ОСНОВЕ БОЛЬШИХ ЯЗЫКОВЫХ МОДЕЛЕЙ

*Статья посвящена актуальной проблеме извлечения ключевых фраз из текстов на естественном языке, что является критически важной задачей в области обработки естественного языка и интеллектуального анализа текста. В ней подробно рассматриваются основные подходы к извлечению ключевых фраз (ключевых слов), включая как традиционные методы, так и современные подходы на основе искусственного интеллекта. В статье рассматривается набор широко используемых методов в этой области, таких как TF-IDF, RAKE, YAKE и методы, основанные на лингвистических анализаторах (парсерах). Эти методы опираются на статистические принципы и графовые структуры, но часто сталкиваются с проблемами, связанными с недостаточной способностью учитывать контекст текста. Большая языковая модель GPT-3 демонстрирует превосходное понимание контекста по сравнению с традиционными методами извлечения ключевых фраз. Эта продвинутая способность позволяет GPT-3 более точно идентифицировать и извлекать релевантные ключевые фразы из текста. Сравнительный анализ с использованием эталонного набора данных Inspec показывает значительно более высокую производительность GPT-3 с точки зрения средней точности (Mean Average Precision, MAP). Однако следует отметить, что, несмотря на высокую точность и качество извлечения, использование больших языковых моделей может быть ограничено в реальном времени из-за их более длительного времени отклика по сравнению с классическими статистическими методами. Таким образом, статья подчеркивает необходимость дальнейших исследований в этой области для оптимизации алгоритмов извлечения ключевых фраз с учетом требований реального времени и контекста текстов.*

*Извлечение ключевых слов; извлечение ключевых фраз; LLM; TF-IDF; большие языковые модели; GPT-3.*

**J. Mohammad**

# KEYPHRASE EXTRACTION BASED ON LARGE LANGUAGE MODELS

*The article addresses the current problem of extracting key phrases from natural language texts, which is a critical task in the field of natural language processing and text mining. It examines in detail the main approaches to extracting key phrases (keywords), including both traditional methods and modern approaches based on artificial intelligence. The paper discusses a set of widely used methods in this field, such as TF-IDF, RAKE, YAKE, and linguistic parser-based methods. These methods are based on statistical principles and/or graph structures, but they often face problems related to their insufficient ability to take into account the context of the text. The GPT-3 large language model demonstrates superior contextual understanding compared to traditional methods for key phrase extraction. This advanced capability allows GPT-3 to more accurately identify and extract relevant key phrases from text. The comparative analysis using the Inspec benchmark dataset reveals GPT-3's significantly higher performance in terms of Mean Average Precision (MAP@K). However, it should be noted that despite high accuracy and extraction quality, the use of large language models may be limited in real-time applications due to their longer response time compared to classical statistical methods. Thus, the article emphasizes the need for further research in this area to optimize key phrase extraction algorithms, taking into account real-time requirements and text context.*

*Keyword extraction; key phrases extraction; LLMs; TF-IDF; Large language models; GPT-3.*

**Введение.** В быстро развивающейся области обработки естественного языка (Natural Language Processing, NLP) извлечение ключевых фраз из текста является важной задачей с широким спектром применений, от классификации текстов до повышения эффективности поисковой оптимизации [1]. Извлечение ключевых фраз включает идентификацию наиболее значимых терминов или выражений в тексте, которые отражают его основные темы и концепции [2].

Традиционные подходы извлечения ключевых фраз включают как контролируемые, так и неконтролируемые методы. Контролируемые методы полагаются на аннотированные наборы данных для обучения моделей, способных идентифицировать ключевые фразы, в то время как неконтролируемые методы, такие как TF-IDF и тематическое моделирование, используют статистические меры для различения важных фраз без необходимости в размеченных данных. Эти подходы были инструментальными во многих приложениях, предлагая баланс простоты и эффективности [3].

Однако появление больших языковых моделей (англ. Large Language Models, LLM) привело к значительным изменениям в области обработки естественного языка. Модели, такие как GPT и BERT, продемонстрировали беспрецедентные возможности в понимании и генерации текста, похожего на человеческий, благодаря своим сложным архитектурам и обширной подготовке на разнообразных наборах данных [4].

LLM используют методы глубокого обучения и огромные объемы данных для захвата сложных моделей языка и контекстуальных нюансов, что делает их исключительно способными извлекать ключевые фразы с высокой точностью. Их способность понимать контекст и семантику превосходит традиционные методы, что позволяет проводить более точные и контекстуально релевантные извлечения.

В поддержку данной теоретической информации в статье представлен сравнительный анализ группы методов извлечения ключевых фраз. Кроме того, их производительность сопоставляется с эффективностью больших языковых моделей.

**Аналитический обзор методов извлечения ключевых фраз.** Методы извлечения ключевых фраз можно условно классифицировать на контролируемые и неконтролируемые, каждый из которых обладает уникальными приемами и областями применения [5, 6]. Ниже, представлено описание основных подходов, а также наиболее значимых методов и инструментов, используемых в данной области.

1. Неконтролируемые подходы. Методы данного подхода не требуют аннотированных обучающих данных и основываются на неотъемлемых свойствах текста. Ключевые методы включают:

*А. Методы, основанные на частоте*. Метод TF-IDF (Term Frequency-Inverse Document Frequency) вычисляет важность термина, учитывая его частоту в документе относительно его частоты в более крупном корпусе. Термины, которые часто встречаются в документе, но редко в других, считаются значимыми. Этот метод эффективен, но может упустить важные фразы, которые встречаются нечасто [7]. Другим примером этих методов являются *YAKE* и *RAKE* [8, 9]. Метод *YAKE* (Yet Another Keyword Extractor) основан на анализе текста без использования стоп-слов[1] (англ. stop-words), что позволяет выделять значимые слова и фразы. Для оценки важности слов *YAKE* применяет комбинацию пяти метрик: нормированная частота, местоположение в тексте, число предложений с выражением, число капитализированных употреблений и сходство со стоп-словами. Это делает его более гибким и точным по сравнению с другими методами, такими как *RAKE*, который оценивает ключевые слова на основе частоты, совместимости и других простых метрик.

*Б. Методы на основе графов*. Эти методы рассматривают документы как узлы в графе и используют алгоритмы для определения центральных или влиятельных узлов, которые соответствуют ключевым фразам. Методы, такие как PageRank и меры центральности, можно использовать для оценки и извлечения наиболее важных фраз. Такие методы, как TextRank, используют графовые структуры для представления связей между словами и фразами. Ранжируя фразы на основе их связности и важности в графе, данный метод может эффективно идентифицировать ключевые фразы [10, 11].

[1] Стоп-слова – это часто встречающиеся слова в языке, такие как "и/and", "в/in", "на/on", "с/with", которые не несут значительной смысловой нагрузки.

*B. Методы на основе вероятности*. Эти методы используют вероятностные модели и статистический анализ для определения ключевых фраз. Скрытое распределение Дирихле (англ. Latent Dirichlet allocation, LDA) – один из самых известных примеров этих методов. Это статистическая модель, которая предполагает, что каждый документ представляет собой смесь различных тем, а каждая тема характеризуется распределением слов [12, 13].

2. Контролируемые подходы. Методы этого подхода используют алгоритмы машинного обучения для классификации фраз на ключевые и неключевые. Эти методы, как правило, обеспечивают более высокую точность, однако требуют значительных затрат времени и ресурсов для подготовки обучающих данных [14].

3. Гибридные подходы. Объединение нескольких методов часто может привести к более точному извлечению ключевых фраз. Например, использование TF-IDF для создания начального списка фраз-кандидатов, а затем применение модели машинного обучения для уточнения и ранжирования этих фраз.

4. Лингвистические методы. Разметка частей речи (англ. Part-of-Speech Tagging, POS) является фундаментальным методом в обработке естественного языка (NLP). Он присваивает грамматические категории словам в тексте, таким как существительные, глаголы, прилагательные и наречия. Этот метод особенно полезен при извлечении ключевых фраз, где цель состоит в том, чтобы определить значимые фразы, которые отражают главные идеи документа [15, 16]. Ниже описаны некоторые приложения POS при извлечении ключевых фраз:

*Идентификация фразы-кандидата*. POS-теги помогают идентифицировать потенциальных кандидатов на роль ключевых фраз, фильтруя слова на основе их грамматических ролей. Например, именные фразы (комбинации прилагательных и существительных) часто являются целевыми, поскольку они обычно представляют ключевые концепции. Применяя определенные шаблоны POS, такие, как «прилагательное + существительное», процесс извлечения может сосредоточиться на более значимых фразах, уменьшая шум от нерелевантных слов.

*Фильтрация нерелевантных слов*. Используя POS-теги, процесс извлечения может исключить слова, которые с меньшей вероятностью будут способствовать значению текста, такие как стоп-слова (например, «and», «the»). Это повышает качество извлекаемых ключевых фраз, гарантируя, что рассматриваются только значимые термины, тем самым улучшая общую релевантность результатов [17].

*Распознавание именованных сущностей (англ. Named Entity Recognition, NER)*. POS-тегирование может помочь в идентификации именованных сущностей, которые часто имеют решающее значение для извлечения ключевой фразы. Помечая слова как собственные имена, организации или местоположения, процесс извлечения может расставить приоритеты для этих сущностей [18].

*Извлечение словосочетаний (англ. Collocation Extraction)*. POS-теги могут использоваться для идентификации словосочетаний – фраз, которые часто встречаются вместе. Анализируя POS-теги соседних слов, система извлечения может распознавать общие фразы, которые могут быть важны для понимания контекста и тем текста [19, 20].

*Улучшение семантической осведомленности*. Недавние подходы интегрировали POS-теги с семантической информацией для повышения производительности методов извлечения ключевых фраз. Учитывая типы слов, извлеченных из POS-тегов, а также контекстно-зависимые семантические критерии, процесс извлечения может создать более релевантные и контекстуально соответствующие ключевые слова.

5. Большие языковые модели (англ. Large Language Models, LLM). Большие языковые модели представляют собой один из самых передовых подходов к извлечению ключевых фраз. GPT-3, разработанная OpenAI, представляет собой современную большую языковую модель, которая использует методы глубокого обучения для понимания и создания текста, похожего на человеческий. GPT-3 использует архитектуру трансформатора, которая использует механизмы внутреннего внимания для обработки входного текста.

Это позволяет модели учитывать контекст слов по отношению друг к другу. Модель GPT-3 предварительно обучена на обширном корпусе текстовых данных, обучаясь предсказывать следующее слово в предложении. Хотя ее можно точно настроить для определенных задач, она также хорошо работает в условиях нулевого или небольшого количества попыток, генерируя выходные данные на основе минимальных примеров или подсказок [4, 21].

Преимущества использования GPT-3 для извлечения ключевых фраз:

♦ Контекстное понимание: способность GPT-3 интерпретировать контекст позволяет извлекать ключевые фразы, которые более релевантны и значимы по сравнению с традиционными методами, полагающимися исключительно на частоту или статистические показатели.

♦ Обучение с нуля: модель может выполнять извлечение ключевых фраз без необходимости использования аннотированных наборах данных. Это делает ее подходящей для приложений, где данных мало или их трудно маркировать.

♦ Гибкость: GPT-3 может адаптироваться к различным доменам и стилям текста, что делает ее универсальной для разных типов документов.

**1. Постановка задачи.** Задачу извлечения ключевых фраз можно сформулировать как задачу ранжирования ключевых фраз-кандидатов по степени их релевантности. Математически это можно описать следующим образом:

Пусть дан текст $\boldsymbol{T}$, состоящий из последовательности слов $\{\mathbf{w_1}, \mathbf{w_2}, \mathbf{..., w_n}\}$. $\boldsymbol{P}$ – множество ключевых фраз-кандидатов, извлеченных из $\boldsymbol{T}$. Введена $\boldsymbol{R(p)}$ – функция ранжирования, которая присваивает оценку каждой ключевой фразе-кандидате $\boldsymbol{p} \in \boldsymbol{P}$. Цель состоит в том, чтобы найти топ-$\boldsymbol{k}$ ключевых фраз из $\boldsymbol{P}$, имеющих наивысшие оценки ранжирования согласно $\boldsymbol{R(p)}$. Математически это можно выразить как задачу оптимизации:

$$\max_{p_1,p_2,\ldots,p_k \in P} \sum_{i=1}^{k} R(p_i), \tag{1}$$

где, $\boldsymbol{p_1, p_2, \ldots, p_k}$ – различные элементы множества $\boldsymbol{P}$, а $\boldsymbol{k}$ – желаемое количество извлекаемых ключевых фраз. Функция ранжирования $\boldsymbol{R(p)}$ может учитывать различные характеристики ключевой фразы-кандидата p, такие как: Частота появления $\boldsymbol{p}$ в $\boldsymbol{T}$; Позиция $\boldsymbol{p}$ в $\boldsymbol{T}$ (например, фразы из заголовков/введения могут иметь больший вес); Наличие стоп-слов в $\boldsymbol{p}$; Семантическая связность/релевантность слов в $\boldsymbol{p}$; Длина $\boldsymbol{p}$ (фразы умеренной длины могут быть предпочтительнее). Эти характеристики могут быть объединены в $\boldsymbol{R(p)}$ с использованием методов линейной регрессии, моделей обучения ранжированию или графовых методов, таких как PageRank.

Использование LLM в этом контексте помогает в вычислении более продвинутых характеристик для $\boldsymbol{R(p)}$, таких как семантическая связность и контекстуальная уместность ключевых фраз, что может привести к улучшению качества ранжирования по сравнению с традиционными статистическими методами.

**2. Вычислительный эксперимент и анализ полученных результатов.** В этом разделе проводится вычислительный эксперимент по оценке эффективности группы методов извлечения ключевых фраз. Данный эксперимент является расширением вычислительного эксперимента, проведенного в предыдущих работах [22, 23], в которых рассматривалась группа таких методов, как TF-IDF, *YAKE*, *RAKE*, $\mathrm{TF_{Spcay}}$, $\mathrm{TF_{stanza}}$ и $\mathrm{TF_{AllenNLP}}$. В данной работе производительность GPT-3 оценивается по сравнению с предыдущими методами при решении задачи извлечения ключевых фраз из документов набора данных *Inspec*.

*Метрика оценки.* Для оценки эффективности методов извлечения ключевых фраз используется модифицированная версия метрики MAP@K:

$$MAP@K = \frac{1}{Q}\sum_{q \in Q} \frac{1}{\min(m,K)} \sum_{i=1}^{K} P(i) \cdot rel(i), \tag{2}$$

где Q – множество ключевых фраз, определенных экспертами.

m – количество ключевых фраз в документе, которые соответствуют ключевым фразам Q.

P(i) – точность на позиции i в ранжированном списке ключевых фраз документа.

rel(i) – бинарная функция сопоставления, равная 1 если ключевая фраза на позиции i соответствует ключевой фразе q, и 0 иначе.

Таким образом, данная метрика позволяет оценить, насколько хорошо система упорядочивает ключевые фразы в документе в соответствии с ключевыми фразами, определенными экспертами.

**3. Реализация и анализ результатов**. Для реализации алгоритма на основе GPT-3 была использована бесплатная версия модели GPT-3, позволяющая отправлять 60 запросов каждый час[2].

Программное приложение было разработано с использованием языка программирования Python для взаимодействия с языковой моделью GPT-3 и получения необходимого ответа. Следующий фрагмент кода иллюстрирует часть реализации.

```python
import pandas as pd
from processing import *
from freeGPT import Client
model = "gpt3"
import json

my_prompt = """Extract the top 20 key phrases from the following text,
                ranking them according to their importance,
                and return the response in JSON format: {key phrases: list} """

def get_kws(my_prompt, doc):
    prompt = my_prompt + doc # doc: документ в обработке

    try:
        response = Client.create_completion(model, prompt) # Sending request
        return response
    except Exception as e:
        print(e)
        return [] # return empty list

```

*Рис. 1. Листинг кода для извлечения ключевой фразы с помощью GPT-3*

Для остальных методов использовалась та же реализация, что и в предыдущей работе автора [23].

В табл. 1 и рис. 2 представлены результаты сравнения различных методов извлечения ключевых фраз с использованием метрики MAP@K, где *K* варьируется от 1 до 20.

GPT-3 демонстрирует впечатляющие результаты в задаче извлечения ключевых фраз. Его производительность, особенно при небольших значениях K, превосходит другие методы, включая $TF_{SpaCy}$ и $TF_{Stanza}$. Максимальное значение MAP@K равное 0.445 при $k$=1 указывает на то, что GPT3 очень эффективно в извлечении наиболее важных ключевых фраз. Даже при увеличении $k$ до 10 и 20 результаты остаются относительно высокими, что свидетельствует о его способности захватывать важные фразы с более высокой точностью по сравнению с другими методами.

Высокая производительность GPT-3 может быть объяснена его архитектурой трансформеров и механизмом внимания. Трансформеры позволяют модели обрабатывать последовательность данных параллельно, что значительно увеличивает ее вычислительную мощность. Механизм внимания позволяет модели сосредоточиться на важных частях входных данных, игнорируя менее значимые детали. Это особенно полезно при извлечении ключевых фраз, так как модель может уделить внимание наиболее релевантным словам или фразам.

[2] freeGPT 1.3.5 https://pypi.org/project/freeGPT/.

Однако следует отметить, что использование GPT-3 может быть ресурсоемким и требовать значительных вычислительных мощностей, особенно при работе с большими наборами данных или в реальных приложениях с ограниченными ресурсами.

Таблица 1

**Точность алгоритмов, измеренная с помощью MAP@K**

| | MAP@K | | | |
|---|---|---|---|---|
| **Методы/ алгоритмы** | @1 | @5 | @10 | @20 |
| TF | 0,18 | 0,096 | 0,058 | 0,046 |
| $TF_{SpaCy}$ | **0,42** | **0,167** | **0,098** | 0,076 |
| $TF_{Stanza}$ | 0,24 | 0,136 | 0,082 | 0,063 |
| $TF_{AllenNLP}$ | 0,25 | 0,13 | 0,08 | 0,06 |
| YAKE | 0,25 | 0,115 | 0,078 | 0,075 |
| RAKE | 0,23 | 0,107 | 0,087 | **0,079** |
| GPT-3 | **0,445** | **0,313** | **0,23** | **0,198** |

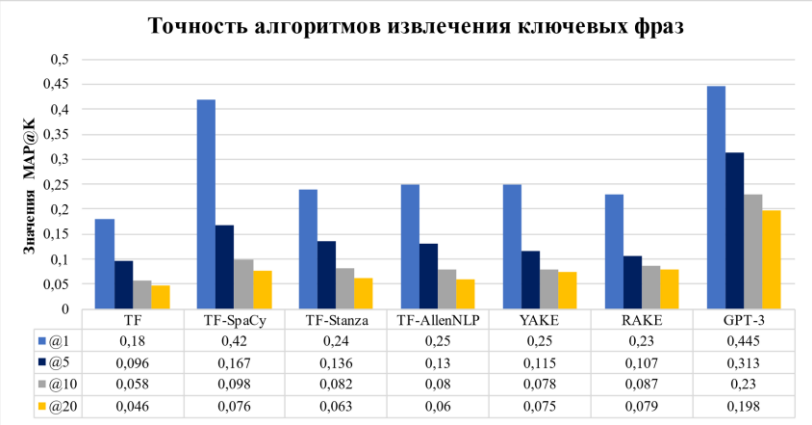



*Рис. 2. Результаты извлечения ключевых фраз из набора данных Inspec, измеренные мерой MAP@K*

Также отмечается, что методы, которые включают в себя дополнительную обработку с помощью лингвистического парсера (библиотек *SpaCy*, *Stanza* и AllenNLP), демонстрируют лучшие результаты по сравнению с чистым TF или другими методами, такими как *YAKE*, *RAKE*. Это указывает на то, что дополнительная семантическая обработка и анализ контекста значительно улучшают точность и качество извлечения ключевых фраз. Этот вариант может быть более предпочтительным по сравнению с GPT-3, при работе в режиме реального времени.

В результате можно сделать вывод, что использование GPT3 для извлечения ключевых фраз является нормальным, учитывая ее способность создавать связный текст и учитывать контекст. Однако, для задач не требующих таких расширенных возможностей, использование традиционных алгоритмов, таких как *YAKE* или *RAKE*, может быть приемлемым, учитывая вычислительные затраты GPT и время ответа на запрос. В таких случаях, традиционные алгоритмы могут быть более эффективными и экономными в ресурсах, несмотря на чуть ниже качество результатов.

**Заключение.** В данной статье рассмотрена важная задача извлечения ключевых фраз из текстов на естественном языке. Анализируются основные подходы к этой проблеме, в том числе традиционные и современные подходы, основанные на больших языковых моделях, в частности модели GPT-3.

Результаты данной работы показали, что традиционные методы, хотя и широко используемые, имеют свои ограничения, особенно в способности учитывать контекст текста. В отличие от них, GPT-3 продемонстрировала значительно лучшие результаты по критерию MAP@K, что подтверждает её эффективность в извлечении ключевых фраз. Тем не менее, важно отметить, что использование больших языковых моделей может быть затруднено в реальном времени из-за более длительного времени отклика по сравнению с классическими статистическими методами.

Таким образом, результаты данной работы подчеркивают необходимость дальнейших исследований и разработок в области оптимизации алгоритмов извлечения ключевых фраз. Это позволит не только повысить точность и качество извлечения, но и сделать эти алгоритмы более подходящими для применения в реальных условиях. В будущем стоит сосредоточиться на разработке гибридных подходов, которые смогут объединить преимущества как традиционных методов, так и современных технологий на основе искусственного интеллекта, чтобы обеспечить более эффективное решение задач извлечения ключевых фраз.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК


1. *Hasan K.S., Ng V.* Automatic keyphrase extraction: A survey of the state of the art // Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics. – 2014. – Vol. 1 – P. 1262-1273.
2. *Schutz A.T.* Keyphrase extraction from single documents in the open domain exploiting linguistic and statistical methods: M. App. Sc Thesis. – 2008.
3. *Mihalcea R., Tarau P.* Textrank: Bringing order into text // Proceedings of the 2004 conference on empirical methods in natural language processing. – 2004. – P. 404-411.
4. *Floridi L., Chiriatti M.* GPT-3: Its Nature, Scope, Limits, and Consequences // Minds and Machines. – 2020. – Vol. 30. GPT-3, No. 4. – P. 681-694.
5. *Kaur J., Gupta V.* Effective approaches for extraction of keywords // International Journal of Computer Science Issues. – 2010. – Vol. 7, No. 6. – P. 144.
6. *Giarelis N., Kanakaris N., Karacapilidis N.* A Comparative Assessment of State-Of-The-Art Methods for Multilingual Unsupervised Keyphrase Extraction // IFIP International Conference on Artificial Intelligence Applications and Innovations. – Springer, 2021. – P. 635-645.
7. *Ramos J.* Using tf-idf to determine word relevance in document queries // Proceedings of the first instructional conference on machine learning. – Citeseer, 2003. – Vol. 242. – P. 29-48.
8. *Rose S., Engel D., Cramer N., Cowley W.* Automatic keyword extraction from individual documents // Text mining: applications theory. – 2010. – Vol. 1. – P. 1-20.
9. *Campos R., Mangaravite V., Pasquali A., Jorge A., Nunes C., Jatowt A.* YAKE! Keyword extraction from single documents using multiple local features // Information Sciences. – 2020. – Vol. 509. – P. 257-289.
10. *Alqaryouti O., Khwileh H., Farouk T., Nabhan A., Shaalan K*. Graph-Based Keyword Extraction // Intelligent Natural Language Processing: Trends and Applications: Studies in Computational Intelligence / eds. K. Shaalan, A.E. Hassanien, F. Tolba. – Cham: Springer International Publishing, 2018. – Vol. 740. – P. 159-172. – ISBN 978-3-319-67055-3.
11. *Beliga S., Meštrović A., Martinčić-Ipšić S.* An overview of graph-based keyword extraction methods and approaches // Journal of information and organizational sciences. – 2015. – Vol. 39, No. 1. – P. 1-20.
12. *Yijun G., Tian X.* Study on keyword extraction with LDA and TextRank combination // Data Analysis and Knowledge Discovery. – 2014. – Vol. 30, No. 7. – P. 41-47.
13. *Cho T., Lee J.-H.* Latent keyphrase extraction using LDA model // Journal of The Korean Institute of Intelligent Systems. – 2015. – Vol. 25, No. 2. – P. 180-185.
14. *Abulaish M., Anwar T.* A supervised learning approach for automatic keyphrase extraction // International Journal of Innovative Computing, Information and Control. – 2012. – Vol. 8, No. 11. – P. 7579-7601.
15. *Akhil K.K., Rajimol R., Anoop V.S.* Parts-of-Speech tagging for Malayalam using deep learning techniques // International Journal of Information Technology. – 2020. – Vol. 12, No. 3. – P. 741-748.
16. *Chiche A., Yitagesu B.* Part of speech tagging: a systematic review of deep learning and machine learning approaches // Journal of Big Data. – 2022. – Vol. 9. – Part of speech tagging. No. 1. – P. 10.
17. *Aro T.O., Dada F., Balogun A.O., Oluwasogo S.A.* Stop words removal on textual data classification. – 2019.
18. *Nadeau D., Sekine S.* A survey of named entity recognition and classification // Lingvisticae Investigationes. – 2007. – Vol. 30, No. 1. – P. 3-26.

19. *Das B., Pal S., Mondal S.K., Dalui D., Shome S.K.* Automatic keyword extraction from any text document using N-gram rigid collocation // Int. J. Soft Comput. Eng. (IJSCE). – 2013. – Vol. 3, No. 2. – P. 238-242.
20. *Evert S., Krenn B.* Exploratory collocation extraction // Phraseology 2005: The Many Faces of Phraseology. – 2005. – P. 113-115.
21. *Maragheh R.Y., Fang C., Irugu C.C., Parikh P., Cho J., Xu J., Sukumar S., Patel M., Korpeoglu E., Kumar S.* LLM-take: theme-aware keyword extraction using large language models // 2023 IEEE International Conference on Big Data (BigData). – IEEE, 2023. LLM-take. – P. 4318-4324.
22. *Мохаммад Ж.Х., Мансур А.М., Кравченко Ю.А., Кравченко Д.Ю.* Метод автоматического извлечения ключевых слов // Международный научно-технический конгресс «Интеллектуальные системы и информационные технологии – 2022». – 2022. – С. 91-97.
23. *Мохаммад Ж.Х., Мансур А.М., Кравченко Ю.А., Бова В.В.* Метод извлечения ключевых фраз на основе новой функции ранжирования // Информационные технологии. – 2022. – Т. 28, № 9. – С. 465-474.

REFERENCES

1. *Hasan K.S., Ng V.* Automatic keyphrase extraction: A survey of the state of the art, *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics,* 2014, Vol. 1, pp. 1262-1273.
2. *Schutz A.T.* Keyphrase extraction from single documents in the open domain exploiting linguistic and statistical methods: M. App. Sc Thesis, 2008.
3. *Mihalcea R., Tarau P.* Textrank: Bringing order into text, *Proceedings of the 2004 conference on empirical methods in natural language processing,* 2004, pp. 404-411.
4. *Floridi L., Chiriatti M.* GPT-3: Its Nature, Scope, Limits, and Consequences, *Minds and Machines,* 2020, Vol. 30. GPT-3, No. 4, pp. 681-694.
5. *Kaur J., Gupta V.* Effective approaches for extraction of keywords, *International Journal of Computer Science Issues,* 2010, Vol. 7, No. 6, pp. 144.
6. *Giarelis N., Kanakaris N., Karacapilidis N.* A Comparative Assessment of State-Of-The-Art Methods for Multilingual Unsupervised Keyphrase Extraction, *IFIP International Conference on Artificial Intelligence Applications and Innovations*. Springer, 2021, pp. 635-645.
7. *Ramos J.* Using tf-idf to determine word relevance in document queries, *Proceedings of the first instructional conference on machine learning*. Citeseer, 2003, Vol. 242, pp. 29-48.
8. *Rose S., Engel D., Cramer N., Cowley W.* Automatic keyword extraction from individual document, *Text mining: applications theory,* 2010, Vol. 1, pp. 1-20.
9. *Campos R., Mangaravite V., Pasquali A., Jorge A., Nunes C., Jatowt A.* YAKE! Keyword extraction from single documents using multiple local features, *Information Sciences,* 2020, Vol. 509, pp. 257-289.
10. *Alqaryouti O., Khwileh H., Farouk T., Nabhan A., Shaalan K*. Graph-Based Keyword Extraction, *Intelligent Natural Language Processing: Trends and Applications: Studies in Computational Intelligence* / eds. K. Shaalan, A.E. Hassanien, F. Tolba. Cham: Springer International Publishing, 2018, Vol. 740, pp. 159-172. ISBN 978-3-319-67055-3.
11. *Beliga S., Meštrović A., Martinčić-Ipšić S.* An overview of graph-based keyword extraction methods and approaches, *Journal of information and organizational sciences,* 2015, Vol. 39, No. 1, pp. 1-20.
12. *Yijun G., Tian X.* Study on keyword extraction with LDA and TextRank combination, *Data Analysis and Knowledge Discovery,* 2014, Vol. 30, No. 7, pp. 41-47.
13. *Cho T., Lee J.-H.* Latent keyphrase extraction using LDA model, *Journal of The Korean Institute of Intelligent Systems,* 2015, Vol. 25, No. 2, pp. 180-185.
14. *Abulaish M., Anwar T.* A supervised learning approach for automatic keyphrase extraction, *International Journal of Innovative Computing, Information and Control,* 2012, Vol. 8, No. 11, pp. 7579-7601.
15. *Akhil K.K., Rajimol R., Anoop V.S.* Parts-of-Speech tagging for Malayalam using deep learning techniques, *International Journal of Information Technology,* 2020, Vol. 12, No. 3, pp. 741-748.
16. *Chiche A., Yitagesu B.* Part of speech tagging: a systematic review of deep learning and machine learning approaches, *Journal of Big Data,* 2022, Vol. 9. Part of speech tagging. No. 1, pp. 10.
17. *Aro T.O., Dada F., Balogun A.O., Oluwasogo S.A.* Stop words removal on textual data classification, 2019.
18. *Nadeau D., Sekine S.* A survey of named entity recognition and classification, *Lingvisticae Investigationes,* 2007, Vol. 30, No. 1, pp. 3-26.
19. *Das B., Pal S., Mondal S.K., Dalui D., Shome S.K.* Automatic keyword extraction from any text document using N-gram rigid collocation, *Int. J. Soft Comput. Eng. (IJSCE),* 2013, Vol. 3, No. 2, pp. 238-242.

20. *Evert S., Krenn B.* Exploratory collocation extraction, *Phraseology 2005: The Many Faces of Phraseology,* 2005, pp. 113-115.
21. *Maragheh R.Y., Fang C., Irugu C.C., Parikh P., Cho J., Xu J., Sukumar S., Patel M., Korpeoglu E., Kumar S.* LLM-take: theme-aware keyword extraction using large language models, *2023 IEEE International Conference on Big Data (BigData).* IEEE, 2023. LLM-take. pp. 4318-4324.
22. *Mokhammad Zh.Kh., Mansur A.M., Kravchenko Yu.A., Kravchenko D.Yu.* Metod avtomaticheskogo izvlecheniya klyuchevykh slov [Method of automatic keyword extraction], *Mezhdunarodnyy nauchno-tekhnicheskiy kongress «Intellektual'nye sistemy i informatsionnye tekhnologii – 2022»* [International scientific and technical congress "Intelligent systems and information technologies - 2022"], 2022, pp. 91-97.
23. *Mokhammad Zh.Kh., Mansur A.M., Kravchenko Yu.A., Bova V.V.* Metod izvlecheniya klyuchevykh fraz na osnove novoy funktsii ranzhirovaniya [Method of key phrase extraction based on a new ranking function], *Informatsionnye tekhnologii* [Information technologies], 2022, Vol. 28, No. 9, pp. 465-474.

Статью рекомендовал к опубликованию к.т.н. С.Г. Буланов.

**Мохаммад Жуман Хуссайн –** Южный федеральный университет; e-mail: zmohammad@sfedu.ru; г. Таганрог, Россия; тел.: 89185433526; кафедра систем автоматизированного проектирования им. В.М. Курейчика; соискатель.

**Mohammad Juman Hussain –** Southern Federal University; e-mail: zmohammad@sfedu.ru; Taganrog, Russia; phone: +79185433526; the Department of Computer-Aided Design Systems named after Viktor Mikhailovich Kureichik; applicant.

# Раздел II. Анализ данных и моделирование

УДК 004.89 DOI 10.18522/2311-3103-2024-5-152-162

**В.И. Волощук, А. Гарягдыев, М.А. Козловская, Я.Э. Мельник, А.Н. Самойлов**

## ПОДХОД К ПОСТРОЕНИЮ АДАПТИВНЫХ СИСТЕМ УЧЕТА ОБЪЕКТОВ С ИСПОЛЬЗОВАНИЕМ МЕТОДОВ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

*Использование методов искусственного интеллекта для учета объектов связано с рядом трудностей, таких как вариативность объектов, влияние условий съемки, перекрытие объектов в сложных сценах, необходимость работы с разными масштабами и высокой точностью, а также наличие шумовых искажений в данных. В статье предлагается основанный на динамическом обучении и адаптации к входным данным подход к организации настройки и эксплуатации адаптивных систем учета объектов на базе методов искусственного интеллекта, включающий в себя несколько последовательных этапов. Первым этапом является семантический анализ запроса пользователя, в основе которого лежит применение векторно-графовой структуры данных, что обеспечивает выделение семантически важных элементов запроса, позволяющий системе понять контекст задачи и адаптировать стратегию поиска и классификации объектов. Далее следует этап автоматического сбора и предобработки данных из открытых источников, что обеспечивает расширение обучающей выборки и повышение устойчивости модели. Следующим важным этапом является формирование обучающей выборки. Этот процесс включает поиск изображений на основе семантики запроса, ручную валидацию и разметку данных, а также первичное обучение системы для автоматической разметки. Выполнение перечисленных этапов повторяется до тех пор, пока не будут достигнуты требуемые характеристики системы. Итеративный процесс дообучения, основанный на чередовании автоматической разметки и ручной корректировки, позволяет сократить временные затраты на формирование обучающих выборок. Преимущество использования векторно-графовой структуры заключается в формировании более точного семантического представления информации. Для повышения обобщающей способности модели применяется аугментация данных, включающая поворот, отражение, масштабирование, изменение яркости и контрастности, а также добавление шума. Предложенный подход предназначен для повышения эффективности (как отношения времени работы системы ко времени её настройки) систем учета объектов, обеспечивая их адаптивность к различным задачам и условиям съемки.*

*Машинное обучение; векторно-графовая структура; обработка изображений; алгоритмы классификации; искусственный интеллект; автоматизация процессов; большие языковые модели; идентификация.*

**V.I. Voloshchuk, A. Garyagdiyev, M.A. Kozlovskaya, Y.E. Melnik, A.N. Samoylov**

## AN APPROACH TO BUILDING ADAPTIVE OBJECT ACCOUNTING SYSTEMS USING ARTIFICIAL INTELLIGENCE METHODS

*The use of artificial intelligence methods for object accounting is associated with a number of difficulties, such as the variability of objects, the influence of shooting conditions, the overlap of objects in complex scenes, the need to work with different scales and high accuracy, as well as the presence of noise distortions in the data. The paper proposes an approach based on dynamic learning and adaptation to input data to organize the setup and operation of adaptive object accounting systems based on artificial intelligence methods, which includes several consecutive stages. The first stage is the semantic analysis of the user's request, which is based on the use of vector-graph data structure, which provides the allocation of semantically important elements of the request, allowing the system to understand the context of the task and adapt the strategy of search and classification of objects. Then follows the stage of automatic*

*collection and preprocessing of data from open sources, which provides the expansion of the training sample and increases the stability of the model. The next important step is the generation of the training sample. This process includes image retrieval based on query semantics, manual validation and data partitioning, and initial training of the system for automatic partitioning. The above steps are repeated until the desired system performance is achieved. The iterative process of pre-training based on alternation of automatic markup and manual correction allows to reduce time expenditures on formation of training samples. The advantage of using vector-graph structure is the formation of more accurate semantic representation of information. Data augmentation including rotation, reflection, scaling, changing brightness and contrast, and adding noise is applied to enhance the generalization ability of the model. The proposed approach is designed to improve the efficiency (as the ratio of system operation time to its setup time) of object registration systems, ensuring their adaptability to different tasks and survey conditions.*

*Machine learning; vector-graph structure; image processing; classification algorithms; artificial intelligence; process automation; large language models; identification.*

**Введение.** Свёрточные нейронные сети (СНС) нашли широкое применение в автоматизации процессов в различных сферах деятельности человека. Однако применение СНС для настройки и эксплуатации адаптивных систем учета объектов сталкивается с рядом трудностей.

Основной трудностью является большая вариативность объектов по форме, размеру, материалу и текстуре, что в свою очередь накладывает ограничение на минимальное количество данных в обучающих выборках [1, 2]. Так, для определения типа покрытия здания на аэроснимке требуется различать металл, бетон, кирпич и дерево, что является сложной задачей классификации. Алгоритм должен уметь выделять уникальные признаки каждого материала, учитывая возможные искажения, вызванные освещением, тенями или углом съемки [3, 4]. Также различные параметры и условия съемки: изменения освещения, тени и рельефа местности, также влияют на характеристики объектов, затрудняя их распознавание. Кроме того, исходные данные часто содержат объекты разного размера, от мелких деталей до крупных сооружений, что требует адаптации алгоритмов СНС к различным масштабам [5–7].

Для повышения эффективности настройки и эксплуатации систем учета объектов требуется снижение временных затрат на подготовку новых обучающих данных, за счет автоматизации процессов поиска данных из открытых источников.

**Предлагаемый подход.** Предлагаемый подход к построению адаптивной системы учета объектов с использованием методов машинного обучения основан на концепции динамического обучения и адаптации к входным данным, учитывая семантику запросов пользователя. На рис. 1 представлена структурная схема, отражающая последовательность действий в рамках предлагаемого подхода.

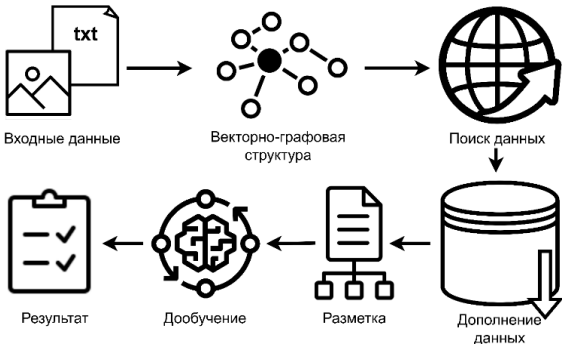



*Рис. 1. Структурная схема предлагаемого подхода*

Подход включает в себя несколько последовательных взаимосвязанных этапов. Сначала при помощи семантического анализа в рамках интерпретации описания единственного исходного изображения извлекается ключевая информация о требуемых объектах. Этот анализ позволяет системе понять контекст запроса и адаптировать стратегию поиска изо-

бражений и классификации на них объектов учёта. Далее следует этап сбора и предобработки данных, предполагающий получение информации из различных открытых источников. После этого осуществляются валидация данных, их очистка от шумов и преобразование в унифицированный формат. Далее производится поиск изображений в открытых источниках, основываясь на семантике, полученной из исходного описания, с дальнейшей итеративной проверкой при помощи эксперта (человек размечает небольшую часть данных). После этого выполняется обучение системы для автоматической разметки оставшихся изображений с возможностью их дальнейшей корректировки человеком.

Такой подход позволяет создать адаптивную систему учета объектов, способную эффективно обрабатывать и обучаться, основываясь на данных из различных источников, а также предоставлять пользователям точные и релевантные результаты. Использование семантического анализа запросов обеспечивает гибкость и адаптивность систем к различным задачам и условиям съемки [8].

**Сравнение с аналогами.** Наиболее распространёнными аналогами предлагаемого подхода в части подбора изображений являются системы поиска снимков по эталонным или по текстовому описанию. Например, Google Images Search и Yandex Image Search. Данные инструменты позволяют найти большое количество изображений из открытых источников, однако обладают недостатком в виде отсутствия возможности одновременного учёта семантики текстовых и визуальных данных запроса пользователя. При использовании предлагаемого подхода удастся избежать этого недостатка путём применения векторно-графовой структуры для обработки текста и применения поиска по схожести цветовой гаммы изображения.

Другим аналогом в части детектирования объектов является модель YOLO-World. Одним из наиболее значимых её преимуществ является ее способность обнаруживать объекты из открытого словаря, что делает ее универсальным инструментом для широкого спектра задач компьютерного зрения. Однако, эта универсальность достигается за счет определенных компромиссов, которые могут снижать точность модели при решении узкоспециализированных задач. Высокая степень обобщения, присущая YOLO-World, обусловлена тем, что модель не обучается на строго ограниченном наборе классов объектов. В то время как это позволяет ей эффективно обнаруживать новые и неизвестные объекты, это также может приводить к снижению точности при работе с хорошо изученными и четко определенными категориями.

Модель, обученная на большом и разнообразном наборе данных, может испытывать трудности в улавливании тонких различий между объектами внутри узкой категории, что приводит к ошибкам классификации и ложным срабатываниям. Кроме того, в условиях открытого словаря существует вероятность смешивания объектов из разных, но визуально похожих категорий, что затрудняет точную классификацию. Архитектура YOLO-World, оптимизированная для работы с большим количеством классов, может быть менее эффективной при решении задач, требующих глубокого анализа изображений и выявления тонких деталей. Предлагаемый в данной статье подход позволит решать узкоспециализированные задачи компьютерного зрения с большей точностью за счёт построения множества адаптивных систем, однако будет требовать большего количества времени из-за необходимости сбора датасета и дообучения существующей модели.

Также данное преимущество будет выражаться в том, что не будет необходимости проектировать всякий раз новую систему, так как каждая из них будет основываться на одной архитектуре. При получении новых наборов данных она позволит после обучения получать систему необходимой специализации.

**Применение векторно-графовой структуры.** Одним из главных преимуществ предлагаемого подхода является применение векторно-графовой структуры [9–11], которая позволяет структурировать информацию семантически более верно. Это помогает производить поиск изображений более точно, основываясь на тексте, им сопутствующем [12]. Рассмотрим пример текстового описания.

***На изображении показан пример фрукта. Это яблоко красного цвета. По его гладкой поверхности видно, что его, вероятно, недавно сорвали с дерева.***

Особенностью векторно-графовой структуры является наличие временной шкалы, которая позволяет рассматривать изменения контекста последовательно. Это помогает рассматривать ситуацию полноценно. Данный текст можно разбить на несколько частей с учётом фактора времени (рис. 2).

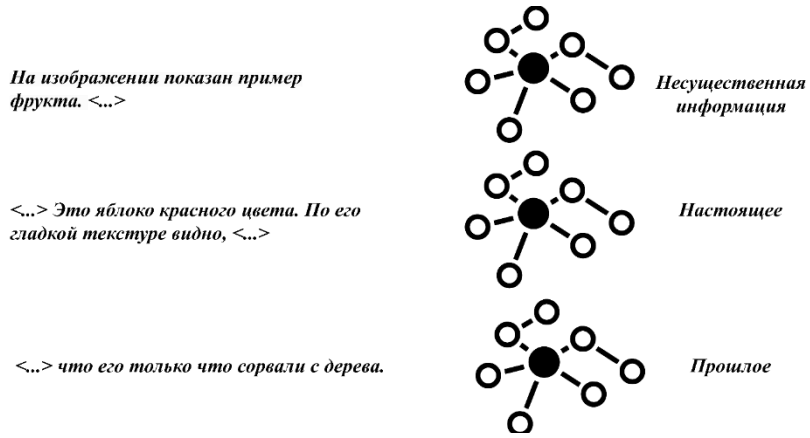



*Рис. 2. Векторно-графовая структура*

На данном рисунке выделены и показаны части текста «настоящее» и «прошлое», которые могут быть полезны при описании статических изображений, содержащих как информацию о ситуации в конкретный момент времени, так и её предысторию и логическое продолжение. Также на указанном рисунке выделена такая часть, как «несущественная информация», наличие которой объясняется тем, что системе заранее известно, что текст представляет собой описание изображения. Стоит отметить, что поиск новых данных будет производиться, основываясь также на тексте, что позволит подбирать изображения более точно. В свою очередь, текст, сопутствующий изображениям в Интернете, может быть различного рода, содержащий в том числе и сложную структуру смены условий сцены.

**Поиск данных.** Для поиска изображений по аналогии и ключевым словам можно использовать подход, представленный на рис. 3. Предполагается проводить предобработку загружаемого пользователем изображения. Изображение будет конвертироваться в стандартный формат RGB, что обеспечит однородность цветовой информации. Затем, с помощью метода перцептивного хеширования (Perceptual Hashing), из изображения будет извлекаться уникальный цифровой отпечаток (хеш), сохраняющий основные визуальные характеристики и устойчивый к изменениям, таким как сжатие или изменение размера.

После выделения основных характеристик изображений было предложено также анализировать вводимую пользователем текстовую информацию. Данное описание будет преобразовываться в ключевые слова, которые будут обрабатываться для удаления стоп-слов и стандартизации. Затем они будут использоваться для выполнения поиска изображений с помощью Google Custom Search API, который сможет возвращать список изображений, релевантных введенному запросу. Каждое найденное изображение будет сопровождаться метаданными, такими как ссылки на источники и описание. Также подход предусматривает поиск изображений по аналогии с загруженным изображением. Этот процесс будет осуществляться через внешний API, поддерживающий поиск по изображениям, и с помощью сравнения перцептивных хешей при помощи метода Hamming distance, измеряющего степень сходства между хешами для идентификации изображения, визуально похожие на эталонное.

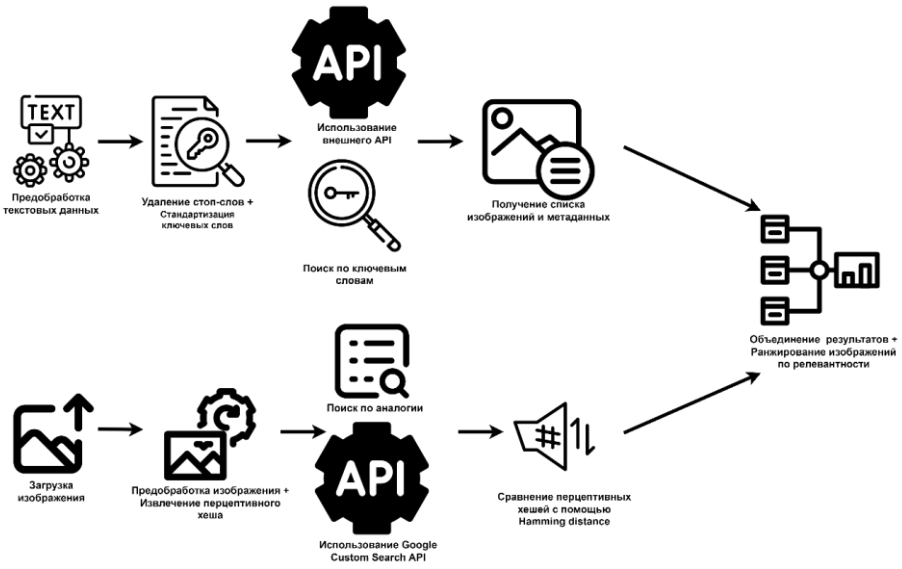



*Рис. 3. Подход поиска данных*

После получения результатов будет производиться их объединение и ранжирование. Затем для каждого изображения будет вычисляться совокупный рейтинг, основанный на релевантности текстового запроса и визуальном сходстве. Результаты будут ранжированы таким образом, чтобы пользователю предлагались изображения, наиболее точно соответствующие запросу пользователя [13–16].

**Разметка данных и обучение системы.** Обучение моделей машинного обучения для построения адаптивной системы учета объектов требует значительного объема данных, чтобы обеспечить высокую точность и устойчивость модели к изменениям в окружающей среде. Однако, в реальных условиях часто возникает проблема недостатка данных или их однотипность. Для решения этой проблемы применяются методы дополнения данных – аугментация.

Процесс дополнения данных начинается с предобработки, которая направлена на приведение данных к единому формату и устранение шума. Стоит отметить, что расширение набора данных с помощью методов аугментации приводит к нескольким положительным эффектам. Во-первых, улучшается обобщающая способность модели, т.е. модель, обученная на более разнообразном наборе данных, лучше справляется с распознаванием объектов в различных условиях, например, при изменении освещения, ракурса или размера объектов. Во-вторых, снижается риск переобучения, который возникает, когда модель слишком хорошо запоминает обучающие данные и не может обобщить знания на новые данные. Дополнение данных помогает избежать переобучения, предоставляя модели больше информации и уменьшая ее зависимость от конкретных обучающих примеров. В-третьих, увеличивается устойчивость модели. Модель, обученная на данных с различными уровнями шума и искажений, становится более устойчивой к шуму и ошибкам в реальных данных.

Для расширения набора данных предлагается применить комплекс методов аугментации. Одним из методов является метод изменения геометрии предполагающий поворот, отражение, масштабирование и сдвиг изображений с объектами. Это позволит модели обучиться распознавать объекты независимо от их ориентации в пространстве, размера и положения на изображении. Второй метод – изменение яркости и контрастности, в результате которого сгенерированные изображения с различными условиями яркости и контрастности, тем самым повысит устойчивость модели к изменениям в освещении.

Также было предложено использовать добавление шума на эталонные данные, что в свою очередь поможет модели стать более устойчивой к изменениям в качестве данных, что особенно важно при работе с реальными изображениями, которые могут содержать шум и искажения при съемке с мобильных устройств.

После проведения операций стандартизации и аугментации данных осуществляется полуавтоматическая разметка новых данных при помощи существующей обученной модели. Этот подход основан на принципе «обучение с учителем» (supervised learning), где модель уже обладает определенными знаниями о распознавании объектов, полученными в процессе первоначального обучения. Предлагаемый в данной работе процесс дообучения модели представлен на рис. 4.

Сначала обученная модель обрабатывает данные, полученные после процесса расширения датасета, выполняя автоматическую разметку объектов интереса. Модель определяет типы объектов и их местоположение на изображении или других типах данных. Полученные результаты автоматической разметки проверяются оператором. Оператор визуально осматривает размеченные данные и в случае ошибок или неточностей вносит необходимые коррективы. Данные, отмеченные оператором как корректные, добавляются к обучающему набору данных и используются для дальнейшего дообучения модели.

Таким образом, система работает в режиме полуавтоматического расширения обучающей выборки, поскольку оператору необходимо вносить коррективы. Однако, несмотря на необходимость наличия человека в данном процессе, предлагаемый подход позволяет сократить время для достижения необходимого результата работы модели в узконаправленной области [17–20].

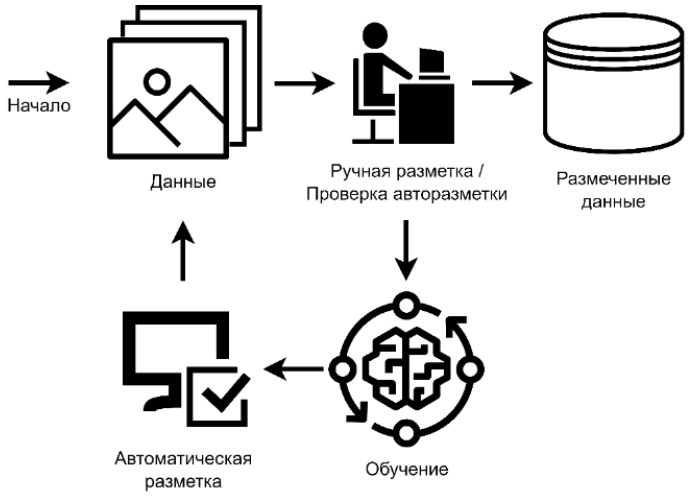



*Рис. 4. Принцип разметки и дообучения системы*

Для оценки влияния аугментации данных на точность моделей распознавания объектов был проведен ряд экспериментов с использованием трех различных архитектур сверточных нейронных сетей: CNN [21, 22], ResNet [23–25] и EfficientNet [26, 27]. Основной целью этих экспериментов было изучение, как методы аугментации, такие как поворот, отражение, масштабирование и изменение яркости изображений, влияют на обобщающую способность и точность моделей.

Результаты экспериментов представлены на рис. 5, который иллюстрирует зависимость точности каждой из моделей от уровня аугментации. Уровень аугментации определяется как доля изображений, к которым были применены указанные методы.

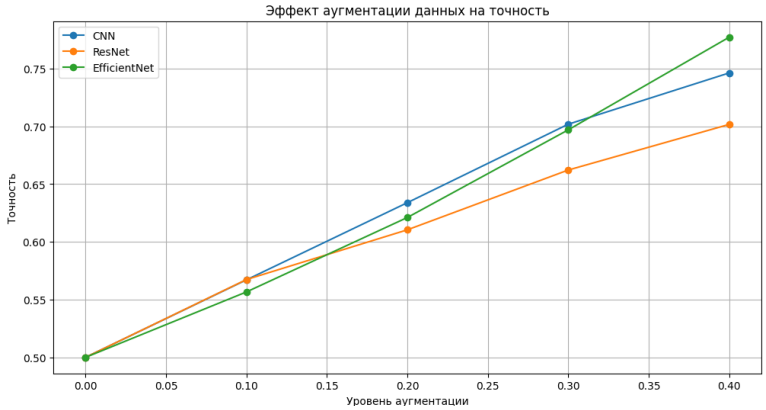



*Рис. 5. Сравнение влияния аугментации данных на точность нейронных сетей*

Согласно полученным данным, наблюдается явное повышение точности всех трех моделей при увеличении уровня аугментации. Этот результат подтверждает эффективность аугментации данных для улучшения обобщающей способности моделей и предотвращения переобучения. Более того, эксперименты показали, что различные архитектуры нейросетей демонстрируют различную чувствительность к аугментации. Например, EfficientNet показывает наибольший прирост точности с увеличением уровня аугментации, в то время как CNN и ResNet показывают менее выраженный эффект.

Важно отметить, что существует оптимальный уровень аугментации, при котором достигается максимальная точность моделей. Дальнейшее увеличение уровня аугментации может приводить к снижению точности, поскольку модель может начать «запоминать» искусственно созданные искажения, а не общие признаки объектов.

На рис. 6 представлена динамика точности каждой модели на обучающей и тестовой выборках в зависимости от количества эпох обучения. Эти результаты играют важную роль в оценке эффективности различных архитектур сверточных нейронных сетей в контексте разработки адаптивных систем учета объектов.

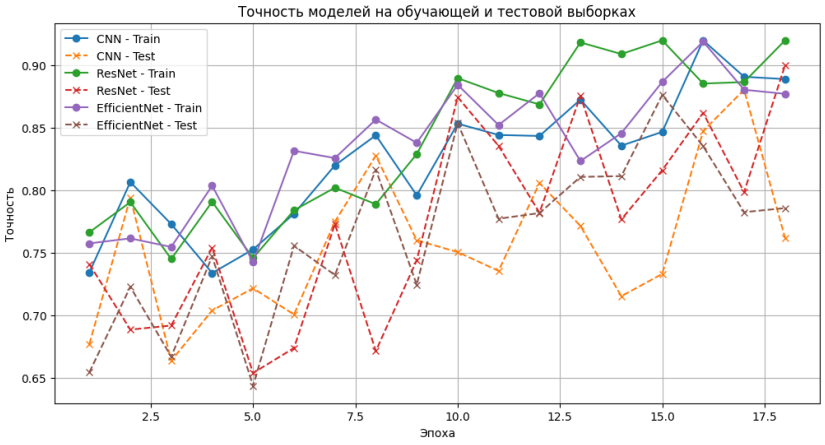



*Рис. 6. Сравнение точности нейронных сетей на обучающей и тестовой выборках*

В ходе эксперимента было выявлено, что разные модели демонстрируют различную скорость обучения. Например, сверточная нейронная сеть CNN быстро достигает высоких результатов на обучающей выборке. Однако, уже после нескольких эпох обучения, она начинает переобучаться, что приводит к снижению точности на тестовой выборке. Это может быть проблемой для систем, работающих с высоко вариативными данными, где важно сохранять способность модели адаптироваться к новым условиям.

В отличие от CNN, архитектуры ResNet и EfficientNet обучаются медленнее, но показывают более стабильные результаты на тестовой выборке. Эти модели имеют особые архитектурные особенности, которые помогают им противостоять переобучению. Например, использование skip connections в ResNet позволяет более эффективно обрабатывать информацию и предотвращает потерю важных признаков, что особенно важно для систем, учитывающих различные объекты на изображениях. EfficientNet, в свою очередь, применяет compound scaling, что помогает оптимально распределять ресурсы и улучшает обобщающую способность модели. Точность на тестовой выборке также продемонстрировала, что EfficientNet показала наилучшие результаты, подтверждая свою высокую обобщающую способность и способность адаптироваться к новым данным.

В результате исследования можно сделать несколько выводов. Во-первых, выбор оптимальной модели нейронной сети должен базироваться на требованиях к точности, скорости обучения и устойчивости к переобучению. Модели, которые обеспечивают высокую производительность на тестовых данных, могут быть более предпочтительными для практического применения. Во-вторых, для задач с ограниченным объемом данных и высокой вариативностью объектов рекомендуется использовать модели с высокой устойчивостью к переобучению, такие как ResNet или EfficientNet.

**Заключение.** В данной работе представлен подход к построению адаптивных систем учета объектов с использованием методов искусственного интеллекта. Ключевыми элементами подхода являются использование СНС для обучения модели учета объектов на изображении, а также применение методов автоматического дополнения данных и полуавтоматической разметки для повышения эффективности систем учета объектов, за счет снижения трудоемкости процессов обучения и разметки данных. Предложенный подход позволяет решать задачи автоматического распознавания объектов и оптимизации процесса учета. Дальнейшие исследования будут направлены на реализацию предложенного в данной работе подхода в рамках адаптивной системы учета объектов.

*Исследование проведено в рамках студенческого научного проекта N 4L/22-04-ПИШ.*

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Казначеева А.А., Захаркина С.В., Власенко О.М., Рыжкова Е.А.* Разработка автоматизированной системы обнаружения дефектов на ткани с применением компьютерного зрения // ИВД. – 2021. – № 12 (84).
2. *Балеев И.А., Земцов А.Н., Зыбин М.И., Смирнов В.А.* Распознавание дефектов на металлических сплавах с помощью алгоритмов компьютерного зрения OpenCV // ИВД. – 2021. – № 3 (75).
3. *Мельник Э.В., Козловский А.В., Онищенко С.В.* Разработка высокопроизводительного метода определения геометрических параметров объектов на изображении // Известия ЮФУ. Технические науки. – 2022. – № 5. – С. 86-97.
4. *Самойлов А.Н., Онищенко С.В., Козловский А.В., Гарягдыев А.М.* Исследование современных подходов и разработка алгоритма для бесконтактного измерения геометрических параметров объектов на цифровом изображении // Системный синтез и прикладная синергетика: Сб. научных работ XI Всероссийской научной конференции. – Ростов-на-Дону – Таганрог. – 2022. – С. 261-266.
5. *Хабарова И.А., Валиев Д.С., Чугунов В.А., Хабаров Д.А.* Современная цифровая фотограмметрия // Международный журнал прикладных наук и технологий «Integral». – 2019. – 4 (2). – С. 41-47.
6. *Краснопевцев Б.В.* Фотограмметрия. – М.: УПП "Репрография" МИИГАиК, 2008. – 160 с.
7. *Алтухов В.Г.* Исследование точности фотограмметрии как метода определения объема объекта // Автоматика и программная инженерия. – 2020. – 2 (32). – С. 69-74.
8. *Обухов А.Д., Патутин К.И., Назарова А.О.* Алгоритмы обработки данных в автоматических системах управления на основе компьютерного зрения // Вестник ТГТУ. – 2022. – № 4.
9. *Kozlovsky A.V., Melnik Y., Voloshchuk V.I.* An Approach for Making a Conversation with an Intelligent Assistant / In: Silhavy, R., Silhavy, P. (eds.) // Artificial Intelligence Application in Networks and Systems. CSOC 2023. Lecture Notes in Networks and Systems. – Vol. 724. – Springer, Cham, 2023. – https://doi.org/10.1007/978-3-031-35314-7_47.
10. *Мельник Я.Э., Лищенко Е.В., Козловская М.А.* О подходе создания алгоритма смены стиля текста на основе его семантического анализа // Технологии разработки инструментальных средств трис-2023: Матер. XIII Международной научно-технической конференции. – Таганрог, 2023. – С. 135-140.

11. *Козловский А.В., Мельник Я.Э., Волощук В.И.* О подходе для автоматической генерации сюжетно связанного текста // Известия Тульского государственного университета. Технические науки. – 2022. – № 9. – С. 160-167.
12. *Шляпников В.М.* Разработка прототипа системы аннотирования изображений для моделей компьютерного зрения // Научные междисциплинарные исследования. – 2020. – № 8-1.
13. *Доморацкий Е.П., Байбикова Т.Н.* Анализ методов поиска изображений в цифровых хранилищах данных // Вестник МФЮА. – 2014. – № 1.
14. *Шаронов А.В., Максимов Н.А., Синча Д.П.* Устойчивый метод поиска изображений в визуальных базах данных // Труды МАИ. – 2011. – № 49.
15. *Бессмертный И.А., Коваль А.А., Белоус Р.О.* Ассоциативный поиск данных с помощью нейронной сети // Научно-технический вестник информационных технологий, механики и оптики. – 2005. – № 19.
16. *Глаговский К.А., Моисеев О.В., Рудельсон Л.Е.* Многокритериальный поиск данных на основе информационных образов // Научный вестник МГТУ ГА. – 2013. – № 9 (195).
17. *Гилязев Р.А., Турдаков Д.Ю.* Активное обучение и краудсорсинг: обзор методов оптимизации разметки данных // Тр. ИСП РАН. – 2018. – № 2.
18. *Бесшапошников Н.О., Кузьменко М.А., Леонов А.Г., Матюшин М.А.* Автоматизация разметки набора данных для нейронных сетей // ВК. – 2018. – № 4 (32).
19. *Михайлов А.А.* Автоматическая разметка данных для сегментации изображений документов с использованием глубоких нейронных сетей // Труды ИСП РАН. – 2022. – № 6.
20. *Якушева С.Ф.* Алгоритм детекции букв и строк на изображениях текстов, набранных печатным шрифтом брайля // Вестник ВГТУ. – 2021. – № 3.
21. *Варшавский П.Р., Кожевников А.В.* Реализация программных средств для классификации данных на основе аппарата сверточных нейронных сетей и прецедентного подхода // Программные продукты и системы. – 2020. – №4.
22. *Андриянов Н.А., Дементьев В.Е., Ташлинский А.Г.* Обнаружение объектов на изображении: от критериев Байеса и Неймана–Пирсона к детекторам на базе нейронных сетей EfficientDet // КО. – 2022. – №1.
23. *Елизаров А.А.* Метод адаптивной классификации изображений с использованием обучения с подкреплением // Программные продукты и системы. – 2022. – № 1.
24. *Авс А., Андриянов Н.А., Соловьев В.И., Соломатин Д.А.* Применение глубокого обучения для аугментации и генерации подводного набора данных с промышленными объектами // Вестник ЮУрГУ. Серия: Компьютерные технологии, управление, радиоэлектроника. – 2023. – № 2.
25. *Кузнецов А.С., Семенов Е.Ю., Матросова Л.Д.* Кластеризация изображений при использовании предобученных нейронных сетей // International Journal of Open Information Technologies. – 2019. – № 7.
26. *Ложкин И.А., Дунаев М.Е., Зайцев К.С., Гармаш А.А.* Аугментация наборов изображений для обучения нейронных сетей при решении задач семантической сегментации // International Journal of Open Information Technologies. – 2023. – №1.
27. *Лихотин М.А.* Использование свёрточных нейронных сетей для анализа изображений // Вестник ВГТУ. – 2023. – №2.

REFERENCES

1. *Kaznacheeva A.A., Zakharkina S.V., Vlasenko O.M., Ryzhkova E.A.* Razrabotka avtomatizirovannoy sistemy obnaruzheniya defektov na tkani s primeneniem komp'yuternogo zreniya [Development of an automated system for the detection of defects on the tissue using computer vision], *IVD* [IVD], 2021, No. 12 (84).
2. *Baleev I.A., Zemtsov A.N., Zybin M.I., Smirnov V.A.* Raspoznavanie defektov na metallicheskikh splavakh s pomoshch'yu algoritmov komp'yuternogo zreniya OpenCV [Defects recognition on the metal alloys with the help of OpenCV computer vision algorithms], *IVD* [IVD], 2021, No. 3 (75).
3. *Mel'nik E.V., Kozlovskiy A.V., Onishchenko S.V.* Razrabotka vysokoproizvoditel'nogo metoda opredeleniya geometricheskikh parametrov ob"ektov na izobrazhenii [Development of a high-performance method for determining the geometric parameters of objects in the image], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2022, No. 5, pp. 86-97.
4. *Samoylov A.N., Onishchenko S.V., Kozlovskiy A.V., Garyagdyev A.M.* Issledovanie sovremennykh podkhodov i razrabotka algoritma dlya beskontaktnogo izmereniya geometricheskikh parametrov ob"ektov na tsifrovom izobrazhenii [Research of modern approaches and development of algorithm for contactless measurement of geometric parameters of objects on digital image], *Sistemnyy sintez i prikladnaya sinergetika: Sb. nauchnykh rabot XI Vserossiyskoy nauchnoy konferentsii* [System synthesis and applied synergetics: Collection of scientific papers of the XI All-Russian scientific conference]. Rostov-on-Don – Taganrog, 2022, pp. 261-266.

5. *Khabarova I.A., Valiev D.S., Chugunov V.A., Khabarov D.A.* Sovremennaya tsifrovaya fotogrammetriya [Modern digital photogrammetry], *Mezhdunarodnyy zhurnal prikladnykh nauk i tekhnologiy «Integral»* [International Journal of Applied Science and Technology "Integral"], 2019, 4 (2), pp. 41-47.
6. *Krasnopevtsev B.V.* Fotogrammetriya [Photogrammetry]. Moscow: UPP "Reprografiya" MIIGAiK, 2008, 160 p.
7. *Altukhov V.G.* Issledovanie tochnosti fotogrammetrii kak metoda opredeleniya ob"ema ob"ekta [Study of the accuracy of photogrammetry as a method for determining the volume of an object], *Avtomatika i programmnaya inzheneriya* [Automation and Software Engineering], 2020, 2 (32), pp. 69-74.
8. *Obukhov A.D., Patutin K.I., Nazarova A.O.* Algoritmy obrabotki dannykh v avtomaticheskikh sistemakh upravleniya na osnove komp'yuternogo zreniya [Data processing algorithms in automatic control systems based on computer vision], *Vestnik TGTU* [TSTU Bulletin], 2022, No. 4.
9. *Kozlovsky A.V., Melnik Y., Voloshchuk V.I.* An Approach for Making a Conversation with an Intelligent Assistant, In: Silhavy, R., Silhavy, P. (eds.), *Artificial Intelligence Application in Networks and Systems. CSOC 2023. Lecture Notes in Networks and Systems,* Vol. 724. Springer, Cham, 2023. Available at: https://doi.org/10.1007/978-3-031-35314-7_47.
10. *Mel'nik Ya.E., Lishchenko E.V., Kozlovskaya M.A.* O podkhode sozdaniya algoritma smeny stilya teksta na osnove ego semanticheskogo analiza [About the approach of creating an algorithm for changing the style of the text on the basis of its semantic analysis], *Tekhnologii razrabotki instrumental'nykh sredstv tris-2023: Mater. XIII Mezhdunarodnoy nauchno-tekhnicheskoy konferentsii* [Technologies for the Development of Instrumental Means TRIS-2023: Proceedings of the XIII International Scientific and Technical Conference]. Taganrog, 2023, pp. 135-140.
11. *Kozlovskiy A.V., Mel'nik Ya.E., Voloshchuk V.I.* O podkhode dlya avtomaticheskoy generatsii syuzhetno svyazannogo teksta [About the approach for automatic generation of plot-related text], *Izvestiya Tul'skogo gosudarstvennogo universiteta. Tekhnicheskie nauki* [News of Tula State University. Technical Sciences], 2022, No. 9, pp. 160-167.
12. *Shlyapnikov V.M.* Razrabotka prototipa sistemy annotirovaniya izobrazheniy dlya modeley komp'yuternogo zreniya [Development of the prototype of the image annotation system for computer vision models], *Nauchnye mezhdistsiplinarnye issledovaniya* [Scientific interdisciplinary research], 2020, No. 8-1.
13. *Domoratskiy E.P., Baybikova T.N.* Analiz metodov poiska izobrazheniy v tsifrovykh khranilishchakh dannykh [Analysis of image search methods in digital data storages], *Vestnik MFYuA* [Vestnik MFSA], 2014, No. 1.
14. *Sharonov A.V., Maksimov N.A., Sincha D.P.* Ustoychivyy metod poiska izobrazheniy v vizual'nykh bazakh dannykh [Stable method of image search in visual databases], *Trudy MAI* [Trudy MAI], 2011, No. 49.
15. *Bessmertnyy I.A., Koval' A.A., Belous R.O.* Assotsiativnyy poisk dannykh s pomoshch'yu neyronnoy seti [Associative data search with the help of neural network], *Nauchno-tekhnicheskiy vestnik informatsionnykh tekhnologiy, mekhaniki i optiki* [Scientific and technical journal of information technologies, mechanics and optics], 2005, No. 19.
16. *Glagovskiy K.A., Moiseev O.V., Rudel'son L.E.* Mnogokriterial'nyy poisk dannykh na osnove informatsionnykh obrazov [Multi-criteria data retrieval based on information images], *Nauchnyy vestnik MGTU GA* [Bulletin of MSTU GA], 2013, No. 9 (195).
17. *Gilyazev R.A., Turdakov D.Yu.* Aktivnoe obuchenie i kraudsorsing: obzor metodov optimizatsii razmetki dannykh [Active learning and crowdsourcing: a review of data markup optimization methods], *Tr. ISP RAN* [Proceedings of ISP RAS], 2018, No. 2.
18. *Besshaposhnikov N.O., Kuz'menko M.A., Leonov A.G., Matyushin M.A.* Avtomatizatsiya razmetki nabora dannykh dlya neyronnykh setey [Automation of data set partitioning for neural networks], *VK* [VK], 2018, No. 4 (32).
19. *Mikhaylov A.A.* Avtomaticheskaya razmetka dannykh dlya segmentatsii izobrazheniy dokumentov s ispol'zovaniem glubokikh neyronnykh setey [Automatic data partitioning for document image segmentation using deep neural networks], *Tr. ISP RAN* [Proceedings of ISP RAS], 2022, No. 6.
20. *Yakusheva S.F.* Algoritm detektsii bukv i strok na izobrazheniyakh tekstov, nabrannykh pechatnym shriftom braylya [Algorithm of letter and line detection on the images of texts typed in printed braille], *Vestnik VGTU* [Bulletin of VSTU], 2021, No. 3.
21. *Varshavskiy P.R., Kozhevnikov A.V.* Realizaciya programmnyh sredstv dlya klassifikacii dannyh na osnove apparata svertochnyh nejronnyh setej i precedentnogo podhoda [Realization of software tools for data classification based on convolutional neural networks and precedent approach], *Programmnye produkty i sistemy* [Software products and systems], 2020, No. 4.

22. *Andriyanov, N.A.; Dementiev, V.E.; Tashlinsky, A.G.* Obnaruzhenie ob"ektov na izobrazhenii: ot kriteriev Bajesa i Nejmana–Pirsona k detektoram na baze nejronnyh setej EfficientDet [Object detection in the image: from Bayes and Neyman-Pearson criteria to detectors based on neural networks EfficientDet] // *KO* [KO], 2022, No. 1.
23. *Elizarov A.A.* Metod adaptivnoj klassifikacii izobrazhenij s ispol'zovaniem obucheniya s podkrepleniem [Method of adaptive image classification using reinforcement learning], *Programmnye produkty i sistemy* [Software Products and Systems], 2022, No. 1.
24. *Avs A., Andriyanov N.A., Soloviev V.I., Solomatin D.A.* Primenenie glubokogo obucheniya dlya augmentatsii i generatsii podvodnogo nabora dannykh s promyshlennymi ob"ektami [Application of deep learning for augmentation and generation of underwater dataset with industrial objects], *Vestnik YuUrGU. Seriya: Komp'yuternye tekhnologii, upravlenie, radioelektronika* [Bulletin SUSU. Series: Computer technologies, management, radio electronics], 2023, No. 2.
25. *Kuznetsov A.S., Semyonov E.Y., Matrosova L.D.* Klasterizatciya izobrazheniy pri ispol'zovanii predobuchennykh nejronnyh setey [Image clustering using pre-trained neural networks], *International Journal of Open Information Technologies* [International Journal of Open Information Technologies], 2019, No. 7.
26. *Lozhkin I.A., Dunayev M.E., Zaitsev K.S., Garmash A.A.* Augmentaciya naborov izobrazhenij dlya obucheniya nejronnyh setej pri reshenii zadach semanticheskoj segmentacii [Augmentation of image sets for training neural networks when solving problems of semantic segmentation], *International Journal of Open Information Technologies,* 2023, No. 1.
27. *Likhotin M.A.* Ispol'zovanie svyortochnykh neyronnykh setey dlya analiza izobrazheniy [Using convolutional neural networks for image analysis], *Vestnik VGTU* [Bulletin VGTU], 2023, No. 2.

Статью рекомендовал к опубликованию д.ф.-м.н., профессор Г.В. Куповых.

**Волощук Вадим Игоревич** – Южный федеральный университет; e-mail: vvoloshchuk@sfedu.ru; г. Таганрог, Россия; тел.: +79081793583; техник-проектировщик Дивизиона «Киберфизические платформы» Передовой инженерной школы «Инженерия киберплатформ».

**Гарягдыев Али** – e-mail: garyagdyev@sfedu.ru; тел.: +79885483343; кафедра вычислительной техники; программист.

**Козловская Мария Алексеевна** – e-mail: arhipenko@sfedu.ru; тел.: +79280681577; инженер Дивизиона «Киберфизические платформы» Передовой инженерной школы «Инженерия киберплатформ».

**Мельник Ярослав Эдуардович** – e-mail: iamelnik@sfedu.ru; тел.: +79885153420; техник-проектировщик Дивизиона «Киберфизические платформы» Передовой инженерной школы «Инженерия киберплатформ».

**Самойлов Алексей Николаевич** – e-mail: asamoylov@sfedu.ru; тел.: +78632184000; кафедра вычислительной техники; зав. кафедрой.

**Voloshchuk Vadim Igorevich** – Southern Federal University; e-mail: vvoloshchuk@sfedu.ru; Taganrog, Russia; phone: +79081793583; design technician of Cyberphysical Platforms Division, Advanced Engineering School Cyber Platforms Engineering.

**Garyagdiyev Ali** – e-mail: garyagdyev@sfedu.ru; phone: +79885483343; the department of Computer Science; programmer.

**Kozlovskaya Maria Alekseevna** – e-mail: arhipenko@sfedu.ru; phone: +79280681577; engineer of Cyberphysical Platforms Division, Advanced Engineering School Cyber Platforms Engineering.

**Melnik Yaroslav Eduardovich** – e-mail: iamelnik@sfedu.ru; phone: +79885153420; design technician of Cyberphysical Platforms Division, Advanced Engineering School Cyber Platforms Engineering.

**Samoylov Aleksey Nikolayevich** – e-mail: asamoylov@sfedu.ru; phone: +78632184000; the Department of Computer Science; head of department.

УДК 004.056 DOI 10.18522/2311-3103-2024-5-163-173

**В.О. Малявина, Е.А. Маро**

# МОДЕЛИРОВАНИЕ УТЕЧЕК ПО ПОБОЧНЫМ КАНАЛАМ ДЛЯ КРИПТОГРАФИЧЕСКОГО АЛГОРИТМОВ «МАГМА» И «КУЗНЕЧИК» НА ОСНОВЕ ЭМУЛЯТОРА ELMO

*Анализ стойкости реализаций средств защиты информации к атакам по побочным каналам является актуальной задачей при разработке криптографических модулей. Первым этапом в исследовании стойкости по побочным каналам рассматривается оценка наличия статистических утечек в различных параметрах работы устройств в ходе выполнения криптографических алгоритмов. Универсальным источником, оцениваемым как побочный канал, рассматривается анализ энергопотребления устройства в ходе криптографических вычислений. В исследовательской работе с помощью инструмента ELMO получены трассы энергопотребления для алгоритмов шифрования «Магма» и «Кузнечик», выявлены инструкции, содержащие статистические утечки по энергопотреблению для исследуемых алгоритмов. Для моделирования трасс энергопотребления в ELMO реализован на языке С алгоритм шифрования ГОСТ Р 34.12—2015 (n=64 «Магма» и n=128 «Кузнечик»). Полнораундовая версия алгоритмов шифрования «Магма» и «Кузнечик» составляет соответственно 15400 инструкций (из них 4450 инструкций содержит потенциальную утечку по энергопотреблению) и 7167 инструкций (из них 4833 инструкций содержит потенциальную утечку по энергопотреблению). Выявление побочного канала (соответствующего обрабатываемым данным) может быть осуществлено с помощью статистического t-теста. Для выполнения этой задачи формируются два независимых набора трасс энергопотребления устройств: трассы при фиксированном значении входных векторов и трассы при произвольных (не совпадающих с фиксированными) значениях входных векторов. Выполнено моделирование утечек по энергопотреблению для различного числа раундов шифрования «Магма» и «Кузнечик» на о снове статистического t-теста. Определены инструкции, содержащие наибольшую статистическую зависимость на базе проведенного тестирования. Для шифра Магма выделены инструкции adds r3,r4,r3 и ldrb r3,[r3,r1], для шифра Кузнечик - lsls r5,r3,#0x0 и str r7,[r3,#0x20000888]. Выявленные инструкции являются оптимальными для последующего проведения дифференциальных или корреляционных атак по энергопотреблению на исследуемые алгоритмы шифрования.*

*Моделирование утечек по энергопотреблению; эмулятор ELMO; симметричный блочный алгоритм шифрования; ГОСТ Р 34.12-2015; шифр «Магма»; шифр «Кузнечик».*

**V.O. Malyavina, E.A. Maro**

# MODELING SIDE-CHANNEL LEAKAGES FOR THE CRYPTOGRAPHIC ALGORITHMS "MAGMA" AND "KUZNACHIK" BASED ON THE ELMO EMULATOR

*Analysis of the resistance of implementations of information security tools to attacks via side channels is a relevant task in the development of cryptographic modules. The first stage in the study of resistance via side channels is the assessment of the presence of statistical leaks in various parameters of the operation of devices during the execution of cryptographic algorithms. The universal source, assessed as a side channel, is the analysis of the energy consumption of the device during cryptographic operations. In this research the ELMO tool was used to obtain power consumption traces for the Magma and Kuznyechik encryption algorithms, identify instructions containing statistical power consumption leaks for observed algorithms. To model the power consumption traces, the GOST R 34.12—2015 encryption algorithm (n=64 Magma and n=128 Kuznyechik) was implemented in C in ELMO. The full-round version of the Magma and Kuznechik encryption algorithms consists of 15,400 instructions (of which 4,450 instructions contain a potential leakage in energy consumption) and 7,167 instructions (of which 4,833 instructions contain a potential leakage in energy consumption), respectively. The side channel (corresponding to the processed data) can be identified using a statistical t-test. To perform this task, two independent sets of device energy consumption traces are formed: traces with a fixed value of the input vectors and traces with arbitrary (not coinciding with the fixed) values of the input vectors. Power consumption leaks were modeled for different numbers of Magma and Kuznyechik encryption rounds based on the statistical t-test. The identified instructions are optimal for subsequent differential or correlation attacks on power consumption on the observed encryption algorithms. The instructions containing the maximal statistical de-*

*pendence based on the conducted testing were determined. For the Magma cipher, the instructions added r3,r4,r3 and ldrb r3,[r3,r1] were identified, for the Kuznyechik cipher - lsls r5,r3,#0x0 and str r7,[r3,#0x20000888]. The identified instructions are optimal for subsequent differential or correlation attacks on power consumption on the encryption algorithms under research.*

*Power consumption leak modeling; ELMO emulator; symmetric block encryption algorithm; GOST R 34.12-2015; Magma cipher; Kuznyechik cipher.*

**Введение.** Классический криптоанализ симметричных шифров рассматривает криптосистему как математический алгоритм, преобразующий некоторый входной текст (или наборы входных текстов) в выходной текст (соответствующий набор выходных текстов) на основе исследователь имеет полное описание преобразований, происходящих внутри криптосистемы, владеет зашифрованными текстами, может обладать соответствующими открытыми текстами (или их частями), но не обладает информацией об используемом секретном ключе. Классические методы криптоанализа опираются на использование недостатков математической конструкции шифра для вычисления ключа шифрования по известным данным, вычислительно быстрее полного перебора множества возможных значений ключей.

На практике криптографический алгоритм не ограничивается только математическим описанием алгоритма шифрования, так как не может существовать без физической реализации в виде конкретного программного или программно-аппаратного средства. Криптографический алгоритм разработан в определенной программной среде, реализуется на определенном оборудовании (типе процессора), что отражается на специфике работы криптосредства и может быть использовано исследователем при криптоанализе.

**Атаки по побочным каналам.** Атаки по побочным каналам представляют собой класс атак, направленный на использование уязвимости (недостатка) в практической реализации криптосистемы. Учитывая важность анализа безопасности различных реализаций криптографических систем, следует отдельно рассматривать стойкость средства защиты информации к атакам по побочным каналам [1–5]. Классификация атак по побочным каналам [6] приведена на рис. 1.

Первоначальным этапом оценки стойкости реализаций криптографических средств защиты к атакам по побочным каналам является выявление утечки, присущей работе криптосистемы. Одним из универсальных каналов утечки для криптографических систем служит канал энергопотребления. Атака по энергопотреблению — пассивная атака, направленная на выявление зависимости между энергопотреблением шифратора (процессора) и преобразуемыми данными с целью получения секретного ключа или защищаемой информации. При проведении атаки по энергопотреблению исследователь должен иметь возможность выполнять измерения энергопотребления с высокой точностью для получения информации о выполняемых на устройстве операциях и их параметрах. Типичная схема стенда для проведения атаки по энергопотреблению показана на рис. 2. Выделяют следующие разновидности атак, в которых используется информация об энергопотреблении: простой анализ энергопотребления [7, 8], дифференциальный анализ энергопотребления [9–11], корреляционный анализа энергопотребления [12–14] и анализ на основе шаблонов [15].

Стойкость реализации к утечкам по энергопотреблению рассматривается как важная составляющая обеспечения заданного уровня безопасности и доверия к средству защиты информации в целом.

В данном исследовании проведено моделирование трасс энергопотребления криптографических средств защиты информации, в основе которых используется реализация алгоритма ГОСТ Р 34.15-2015 [16] (n=64 «Магма»), и выполнен поиск наличия каналов утечки, путем выявления наборов инструкций, для которых имеются статистические зависимости энергопотребления устройства от значения обрабатываемых данных (по результатам t-теста).

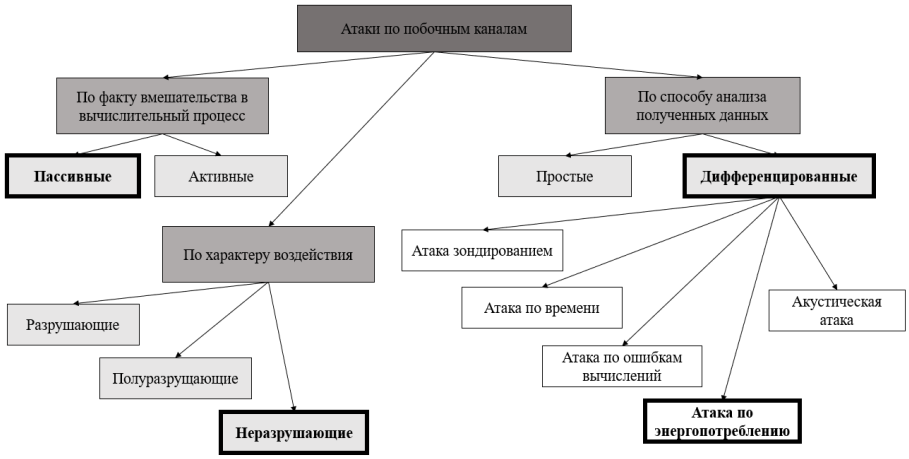



*Рис. 1. Классификация атак по побочным каналам.*

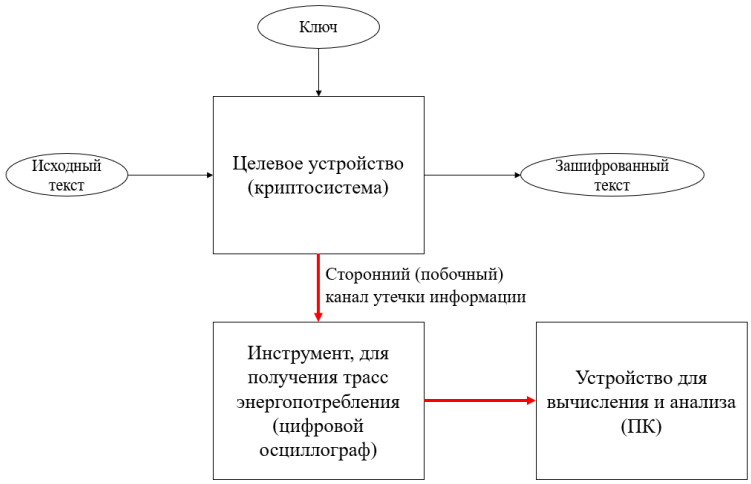



*Рис. 2. Структура стенда для проведения атаки по энергопотреблению.*

**Моделирование энергопотребления с помощью инструмента ELMO.** Любое моделирование энергопотребления (или другого побочного канала) состоит из двух составных частей: эмуляция процесса, который выполняется внутри устройства, и моделирование наблюдаемого извне поведения устройства (для сопоставления эмулируемых процессов с прогнозируемым потреблением). Моделирование можно в общих чертах разделить на категории в зависимости от архитектурного уровня, на котором они пытаются охарактеризовать мощность:

1. Моделирование транзисторного уровня. При наличии достаточной информации о технологии, по которой будет построен чип, схему можно сопоставить с сетью транзисторов, потребляемая мощность которых моделируется с помощью известных дифференциальных уравнений.

2. Моделирование уровня шлюза. Данный вид также основан на списках соединений (с обратной аннотацией). Для моделирования количество переходов в каждом шлюзе подсчитывается и взвешивается в соответствии с информацией в списке соединений. Тогда сумма по всем взвешенным переходам является приближением мгновенной мощности схемы.

3. Моделирование поведенческого уровня. На этом уровне нет информации о размещении элементов схемы и маршрутизации сигналов между ними. Доступ имеется только к поведенческому описанию компонентов – например, в форме машинного кода или микрокода низкого уровня, кода инструкции/ассемблера или кода более высокого уровня (например, C). Разработка точных моделей на этом уровне требует доступа к реальным устройствам (и лабораторной установке), с помощью которых оценивается средняя мощность различных (последовательностей) инструкций. Они хорошо подходят для небольших устройств (и, следовательно, низкой сложности), в которых инструкции по сборке сопоставляются непосредственно с машинными инструкциями без дальнейшего декодирования в микроинструкции.

ELMO [17] – это инструмент моделирования, совмещающий несколько категорий из описанных выше. Инструмент ELMO содержит в себе эмулятор конкретной архитектуры (Arm Cortex-M0) с эмпирически оцененными моделями энергопотребления, зависящими от входных данных.

Программное средство ELMO основано на эмуляторе инструкций для Thumb под названием Thumbulator. Thumbulator предназначен для воспроизведения работы процессоров семейства ARM Cortex M0 при выполнении симметричных блочных шифров. Thumbulator принимает на вход двоичную программу на ассемблере Thumb и транслирует ее в машинные инструкции, что позволяет воспроизвести поток данных ядра микропроцессора с достаточной точностью [18].

Для моделирования энергопотребления криптосистем, основанных на ARM Cortex M0, в ELMO интегрированы широко используемые в реализациях симметричной криптографии модели инструкции Thumb. Эти инструкции можно распределить на 5 разных групп:

1) инструкции загрузки (ldr, ldrb, ldrh);
2) инструкции ALU (adds, adds #imm, ands, eors, movs, movs #imm, orrs, subs, subs #imm, cmp, cmp #imm);
3) инструкции сохранения (str, strb, strh);
4) инструкции сдвига (lsls, lsrs, rors);
5) инструкция умножения (muls).

Инструкции с утечкой получены в результате выполнения статистического поиска на основе t-теста в ELMO. Значение параметра t вычисляется по формуле (1):

$$t = \frac{mean_{fix} - mean_{rand}}{\sqrt{\frac{var_{fix}}{N_{fix}} + \frac{var_{rand}}{N_{rand}}}}, \quad (1)$$

где *fix* – группа фиксированных значений,
*rand* – группа случайных значений,
*mean* – среднее значение всех трасс в группе,
*var* – стандартное отклонение выборки всех трасс в группе,
*N* – размер группы.

Пороговым значением, по которому делается вывод о наличии утечки, зафиксировано значение t = |4.5|, в соответствии с рекомендациями стандарта CSA ISO/IEC 17825-2018 "Information technology - Security techniques - Testing methods for the mitigation of non-invasive attack classes against cryptographic modules" [19] по использованию Test Vector Leakage Assessment (TVLA) [20, 21].

**Моделирование энергопотребления шифра «Магма».** В алгоритме «Магма» для шифрования используется блок размером 64 бита, длина ключа составляет 256 бит. Раундовые ключи получаются из исходного путем его деления на восемь 32-битных подключей (Ki). После получения подключей идет непосредственно процесс шифрования: блок входных данных разделяется на две равные по длине части – правую (R) и левую (L) (по 32 бита каждая), над которыми выполняется тридцать две итерации раундового преобразования с использованием раундовых ключей. На рис. 3 представлена схема раундового преобразования алгоритма «Магма» при шифровании.

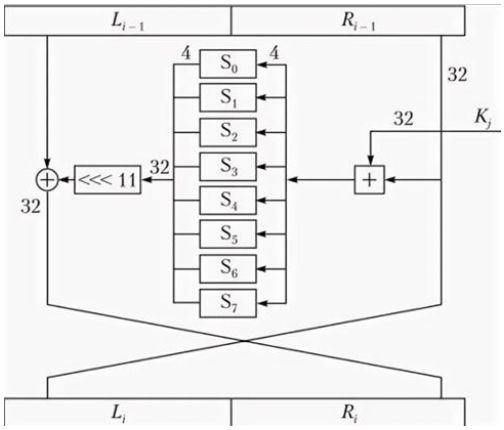



*Рис. 3. Схема раундового преобразования алгоритма «Магма»*

Проведена оценка общего количества инструкций и количества инструкций, содержащих статистические утечки по энергопотреблению, для различного числа раундов. Результаты моделирования представлены в табл. 1.

На рис. 4 представлена одна трасса энергопотребления полнораундового шифра «Магма». Из данного графика видно, как сначала выполняются служебные инструкции инициализации параметров шифрования (независимые от количества раундов), после чего запускаются группы инструкций раундового преобразования (соответствующие 32 пика).

Таблица 1

**Результаты моделирования утечек по побочным каналам для различного количества раундов шифра «Магма»**

| Кол-во раундов | Общее кол-во инструкций | Инструкции с утечками по энергопотреблению | Процентное соотношение |
|---|---|---|---|
| 32 | 15400 | 4450 | 29% |
| 16 | 8820 | 2203 | 25% |
| 2 | 2992 | 462 | 15% |
| 1 | 2599 | 277 | 11% |

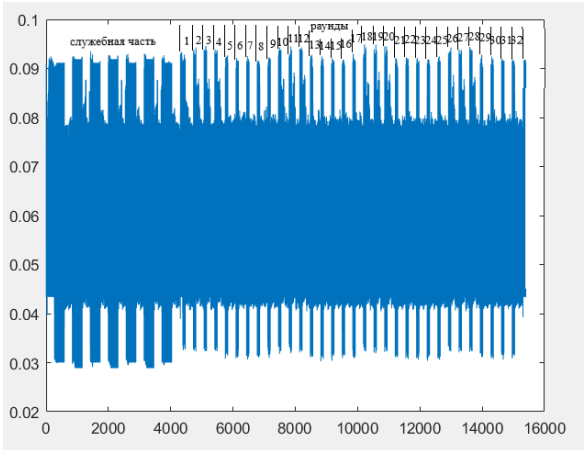



*Рис. 4. Сгенерированная с помощью инструмента ELMO трасса энергопотребления при выполнении тридцати двух раундов алгоритма «Магма»*

На рис. 5 представлены результаты теста FixedvsRandom для одного раунда шифра «Магма». Наиболее уязвимые для атак посторонним каналам инструкции занесены в табл. 2, курсивом выделены два наибольших значения по результатам статистического теста.

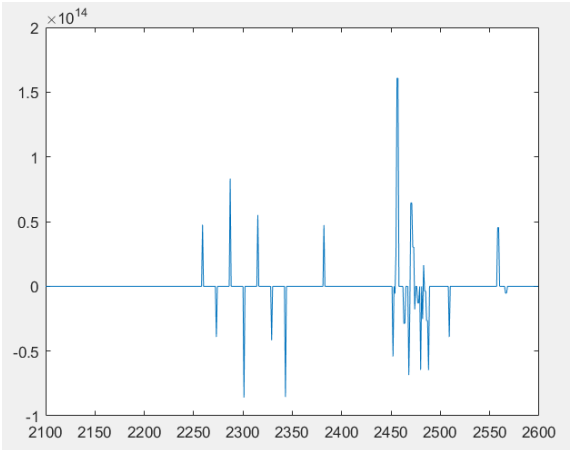



*Рис. 5. Результаты применения статистического теста FixedvsRandom в эмуляторе ELMO для одного раунда шифра «Магмы»*

Таблица 2

**Инструкции, содержащие статистическую зависимость (утечку), для одного раунда шифра «Магма»**

| Номер инструкции | Адрес | Машинный код | Ассемблерный код инструкции |
|---|---|---|---|
| 2287 | 0x0800014A | 0x090B | lsrs r3,r1,#0x4 |
| 2301 | 0x0800014A | 0x090B | lsrs r3,r1,#0x4 |
| 2343 | 0x0800014A | 0x090B | lsrs r3,r1,#0x4 |
| *2456* | *0x08000176* | *0x18E3* | *adds r3,r4,r3* |
| *2457* | *0x08000178* | *0x5C5B* | *ldrb r3,[r3,r1]* |
| 2468 | 0x08000172 | 0x011B | lsls r3,r3,#0x4 |
| 2470 | 0x08000176 | 0x18E3 | adds r3,r4,r3 |
| 2471 | 0x08000178 | 0x5C5B | ldrb r3,[r3,r1] |
| 2480 | 0x0800016E | 0x090B | lsrs r3,r1,#0x4 |
| 2488 | 0x0800017A | 0x5483 | strb r3,[r0,r2] |

**Моделирование энергопотребления шифра «Кузнечик».** Шифр «Кузнечик» представляет собой симметричный блочный алгоритм, работающий с блоками данных длинной 128 бит. Размер ключа составляет 256 бит. Основу алгоритма составляет подстановочно-перестановочная сеть (SP-сеть). Алгоритм шифрования «Кузнечик» состоит из выполнения девяти полных раундов, каждый из которых включает в себя три последовательные операции. Первая операция представляет собой сложение по модулю 2 (XOR) ключа и входного блока данных, вторая операция производит нелинейное преобразование, которое представляет собой простую замену одного байта на другой в соответствии с таблицей, третья операция называется линейным преобразованием, при котором каждый байт из блока умножается в поле Галуа на один из коэффициентов ряда (148, 32, 133, 16, 194, 192, 1, 251, 1, 192, 194, 16, 133, 32, 148, 1) в зависимости от порядкового номера байта. Затем байты складываются между собой по модулю 2, и весь блок сдвигается в сторону младшего разряда, а полученное число записывается на место считанного байта. Последний десятый раунд является не полным и включает в себя только операцию сложения с ключом.

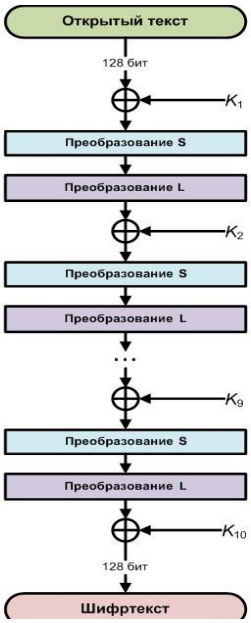



*Рис. 6. Схема шифрования алгоритма «Кузнечик»*

Аналогично проведенному исследованию утечек по энергопотреблению для шифра «Магма» было вычислено общее количество инструкций для разного количества раундов шифра «Кузнечик» и с помощью статистического t-теста в ELMO выявлены инструкции, содержащие утечку по энергопотреблению. Результаты моделирования трасс энергопотребления шифра «Кузнечик» и исследования инструкций с утечками занесены в табл. 3.

Таблица 3

**Результаты моделирования утечек по побочным каналам для различного количества раундов шифра «Кузнечик»**

| Кол-во раундов | Общее кол-во инструкций | Инструкции с утечками по энергопотреблению | Процентное соотношение |
|---|---|---|---|
| 10 | 7167 | 4833 | 67% |
| 5 | 3735 | 2492 | 67% |
| 2 | 1674 | 1084 | 65% |
| 1 | 970 | 611 | 63% |

Трасса энергопотребления полнораундового шифра «Кузнечик» показана на рис. 7. На данном графическом представлении визуально выделяются 10 раундов шифрования.

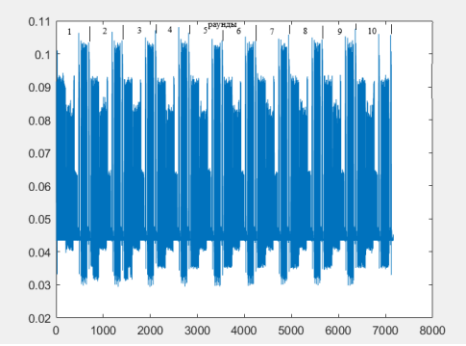



*Рис. 7. Сгенерированная с помощью инструмента ELMO трасса энергопотребления при выполнении десяти раундов алгоритма «Кузнечик»*

Проведен анализ на базе теста FixedvsRandom для одного раунда шифра «Кузнечик», результаты в графическом виде приведен на рис. 8. В табл. 4 занесены десять инструкций, содержащих выявленные статические утечки по энергопотреблению. Курсивом выделены инструкции, с максимальным отклонением по результатам статистического теста.

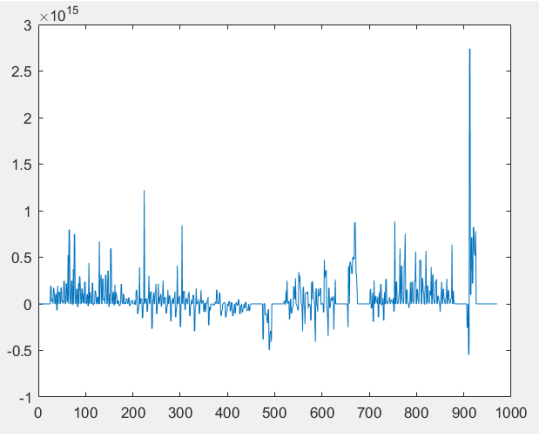



*Рис. 8. Результаты применения теста FixedvsRandom в эмуляторе ELMO для одного раунда шифра «Кузнечик»*

Таблица 4

Инструкции, содержащие статистическую утечку по энергопотреблению для одного раунда шифра «Кузнечик»

| Номер инструкции | Адрес | Машинный код | Ассемблерный код инструкции |
|---|---|---|---|
| 65 | 0x08000192 | 0x545A | strb r2,[r3,r1] |
| 224 | 0x080001AA | 0x3301 | adds r3,#0x01 |
| 304 | 0x080001AA | 0x3301 | adds r3,#0x01 |
| 669 | 0x080009CA | 0x609F str | r7,[r3,#0x20001EA8] |
| 754 | 0x0800025C | 0x702A | strb r2,[r5,#0x20001EA4] |
| *912* | *0x080009C0* | *0x001D* | *lsls r5,r3,#0x0* |
| *913* | *0x080009C2* | *0x601F* | *str r7,[r3,#0x20000888]* |
| 920 | 0x080009C8 | 0x68A7 | ldr r7,[r4,#0x20001EA8] |
| 921 | 0x080009CA | 0x609F | str r7,[r3,#0x20000890] |
| 925 | 0x080009CE | 0x3410 | adds r4,#0x10 |

**Заключение.** В рамках исследования проведено моделирование трасс энергопотребления для различного количества раундов алгоритмов шифрования «Магма» и «Кузнечик». Выделены сигнатуры трасс энергопотребления, соответствующие отдельным раундам и преобразованиям шифрования исследуемого алгоритма. Выполнены наборы статистических тестов (t-тестов), по которым определены инструкции, содержащие утечки для одного раунда шифров «Магма» и «Кузнечик». Сформированные файлы трасс энергопотребления, ассемблерного кода и FxedvsRandom тестов предоставлены в общий доступ для возможности ознакомления и последующего использования результатов моделирования.

Полнораундовая версия алгоритма шифрования «Магма» содержит 15400 инструкций, из них 4450 инструкций содержит потенциальную утечку по энергопотреблению. Для одного раунда шифра «Магма» инструкции с номерами 2456 (adds r3,r4,r3) и 2457 (ldrb r3,[r3,r1]) имеют максимальное значение по статистическому тесту FixedvsRandom.

Полнораундовая версия алгоритма шифрования «Кузнечик» содержит 7167 инструкций, из них 4833 инструкций содержит потенциальную утечку по энергопотреблению. Для одного раунда шифра «Кузнечик» инструкции с номерами 912 (lsls r5,r3,#0x0) и 913 (str r7,[r3,#0x20000888]) имеют максимальное значение по статистическому тесту FixedvsRandom.

Выявленные инструкции и, следовательно, точки на трассах энергопотребления являются оптимальными для проведения анализа стойкости исследуемых реализаций алгоритмов шифрования к атаке по побочному каналу энергопотребления.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК


1. *Hou X., Breier J.* Side-Channel Analysis Attacks and Countermeasures // Cryptography and Embedded Systems Security. Springer, Cham. – https://doi.org/10.1007/978-3-031-62205-2_4.
2. *Piessens F. and van Oorschot P.C*. Side-Channel Attacks: A Short Tour // IEEE Security & Privacy. – March-April 2024. – Vol. 22, No. 2. – P. 75-80. – DOI: 10.1109/MSEC.2024.3352848.
3. *Kaleem M., Mushtaq M., Ali Ramay S., Aamir Mahmood, Abbas Khan T., Kamran Hussain S., Anwar A., Abdullah Bhatti H.* Navigating Side-Channel Attacks: A Comprehensive Overview of Cryptographic System Vulnerabilities // Journal of Computing & Biomedical Informatics. – 2024. – 7 (02). – https://jcbi.org/index.php/Main/article/view/626.
4. *Cui X., Zhang H., Xu J., Fang X., Ning W., Wang Y., Hosen M.S.* A Data Augmentation Method for Side-Channel Attacks on Cryptographic Integrated Circuits // Electronics. – 2024. – 13. – 1348. – https://doi.org/10.3390/electronics13071348.
5. *Amrouche A., Boubchir L. and Yahiaoui S.* Side Channel Attack using Machine Learning // 2022 Ninth International Conference on Software Defined Systems (SDS), Paris, France, 2022. – P. 1-5. – DOI: 10.1109/SDS57574.2022.10062906.
6. *Krasovsky A.V. and Maro E.A*. Actual and historical state of side channel attacks theory // Proceedings of the 12th International Conference on Security of Information and Networks (SIN '19). Association for Computing Machinery, New York, NY, USA. – Article 13. – P. 1-7. – https://doi.org/10.1145/3357613.3357627
7. *Kitazawa T., Fujimoto D. and Hayashi Y.* Fundamental Study on Simple Power Analysis Using Backscattering from Switching Regulators // 2024 International Symposium on Electromagnetic Compatibility – EMC Europe, Brugge, Belgium, 2024. – P. 22-26. – DOI: 10.1109/EMCEurope59828.2024.10722404.
8. *Camacho-Ruiz E., Sánchez-Solano S., Martínez-Rodríguez M.C., Tena-Sanchez E. and Brox P.* A Simple Power Analysis of an FPGA implementation of a polynomial multiplier for the NTRU cryptosystem // 2023 38th Conference on Design of Circuits and Integrated Systems (DCIS), Málaga, Spain, 2023. – P. 1-6. – DOI: 10.1109/DCIS58620.2023.10336001.
9. *Xu J., Fan A., Lu M. and Shan W*. Differential Power Analysis of 8-Bit Datapath AES for IoT Applications // 2018 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications / 12th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE), New York, NY, USA, 2018. – P. 1470-1473. – DOI: 10.1109/TrustCom/BigDataSE.2018.00205.
10. *Wang W., Yu Y., Standaert F.-X., Liu J., Guo Z. and Gu D*. Ridge-Based DPA: Improvement of Differential Power Analysis For Nanoscale Chips // IEEE Transactions on Information Forensics and Security. – May 2018. – Vol. 13, No. 5. – P. 1301-1316. – DOI: 10.1109/TIFS.2017.2787985.
11. *Cai X., Li R., Kuang S., Tan J.* An Energy Trace Compression Method for Differential Power Analysis Attack // IEEE Access. – 2020. – Vol. 8. – P. 89084-89092.
12. *Fernandes Medeiros S., Gérard F., Veshchikov N., Lerman L., Markowitch O.* Breaking Kalyna 128/128 with Power Attacks / in Carlet, C., Hasan, M., Saraswat, V. (eds) // Security, Privacy, and Applied Cryptography Engineering. SPACE 2016. Lecture Notes in Computer Science. – Vol. 10076. – Springer, Cham. – https://doi.org/10.1007/978-3-319-49445-6_23.
13. *Jeon Y., Yoon J.W.* Filtering-Based Correlation Power Analysis (CPA) with Signal Envelopes Against Shuffling Methods / You I. (eds) // Information Security Applications. WISA 2020. Lecture Notes in Computer Science. – Vol. 12583. – Springer, Cham. – https://doi.org/10.1007/978-3-030-65299-9_29.
14. *Lo O., Buchanan W.J., Carson D.* Power analysis attacks on the AES-128 S-box using differential power analysis (DPA) and correlation power analysis (CPA) // Journal of Cyber Security Technology. – 2016. – 1 (2). – P. 88-107. – https://doi.org/10.1080/23742917.2016.1231523.
15. *Xin J., Du Z.* Template attack based on uBlock cipher algorithm // Frontiers in Computing and Intelligent Systems. – 2023. – 3 (1). – P. 90-93. – https://doi.org/10.54097/fcis.v3i1.6031.
16. ГОСТ Р 34.12-2015 Информационная технология. Криптографическая защита информации. Блочные шифры. – URL: https://tc26.ru/standard/gost/GOST_R_3412-2015.pdf.
17. Statistical leakage simulator for the ARM M0 family ELMO. – URL: https://github.com/scaresearch/ELMO.

18. *Welch D.* Thumbulator. – URL: https://github.com/dwelch67/thumbulator.git.
19. CSA ISO/IEC 17825-2018 Information technology - Security techniques - Testing methods for the mitigation of non-invasive attack classes against cryptographic modules.
20. *Goodwill G., Jun B., Jaffe J. and Rohatgi P*. A testing methodology for side-channel resistance validation // NIST Non-Invasive At-tack Testing Workshop. – 2011.
21. *Cooper J., DeMulder E., Goodwill G., Jaffe J., Kenworthy G. and Rohatgi P.* Test vector leakage assessment (tvla) methodology in practice // International Cryptographic Module Conference. – 2013.

REFERENCES

1. *Hou X., Breier J.* Side-Channel Analysis Attacks and Countermeasures, *Cryptography and Embedded Systems Security. Springer, Cham*. Available at: https://doi.org/10.1007/978-3-031-62205-2_4.
2. *Piessens F. and van Oorschot P.C*. Side-Channel Attacks: A Short Tour, *IEEE Security & Privacy,* March-April 2024, Vol. 22, No. 2, pp. 75-80. DOI: 10.1109/MSEC.2024.3352848.
3. *Kaleem M., Mushtaq M., Ali Ramay S., Aamir Mahmood, Abbas Khan T., Kamran Hussain S., Anwar A., Abdullah Bhatti H.* Navigating Side-Channel Attacks: A Comprehensive Overview of Cryptographic System Vulnerabilities, *Journal of Computing & Biomedical Informatics,* 2024, 7 (02). Available at: https://jcbi.org/index.php/Main/article/view/626.
4. *Cui X., Zhang H., Xu J., Fang X., Ning W., Wang Y., Hosen M.S.* A Data Augmentation Method for Side-Channel Attacks on Cryptographic Integrated Circuits, *Electronics,* 2024, 13, 1348. Available at: https://doi.org/10.3390/electronics13071348.
5. *Amrouche A., Boubchir L. and Yahiaoui S.* Side Channel Attack using Machine Learning, *2022 Ninth International Conference on Software Defined Systems (SDS), Paris, France, 2022,* pp. 1-5. DOI: 10.1109/SDS57574.2022.10062906.
6. *Krasovsky A.V. and Maro E.A*. Actual and historical state of side channel attacks theory, *Proceedings of the 12th International Conference on Security of Information and Networks (SIN '19). Association for Computing Machinery, New York, NY, USA,* Article 13, pp. 1-7. Available at: https://doi.org/10.1145/3357613.3357627
7. *Kitazawa T., Fujimoto D. and Hayashi Y.* Fundamental Study on Simple Power Analysis Using Backscattering from Switching Regulators, *2024 International Symposium on Electromagnetic Compatibility – EMC Europe, Brugge, Belgium, 2024,* pp. 22-26. DOI: 10.1109/EMCEurope59828.2024.10722404.
8. *Camacho-Ruiz E., Sánchez-Solano S., Martínez-Rodríguez M.C., Tena-Sanchez E. and Brox P.* A Simple Power Analysis of an FPGA implementation of a polynomial multiplier for the NTRU cryptosystem, *2023 38th Conference on Design of Circuits and Integrated Systems (DCIS), Málaga, Spain, 2023,* pp. 1-6. DOI: 10.1109/DCIS58620.2023.10336001.
9. *Xu J., Fan A., Lu M. and Shan W*. Differential Power Analysis of 8-Bit Datapath AES for IoT Applications, *2018 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications / 12th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE), New York, NY, USA, 2018,* pp. 1470-1473. DOI: 10.1109/TrustCom/BigDataSE.2018.00205.
10. *Wang W., Yu Y., Standaert F.-X., Liu J., Guo Z. and Gu D*. Ridge-Based DPA: Improvement of Differential Power Analysis For Nanoscale Chips, *IEEE Transactions on Information Forensics and Security,* May 2018, Vol. 13, No. 5, pp. 1301-1316. DOI: 10.1109/TIFS.2017.2787985.
11. *Cai X., Li R., Kuang S., Tan J.* An Energy Trace Compression Method for Differential Power Analysis Attack, *IEEE Access,* 2020, Vol. 8, pp. 89084-89092.
12. *Fernandes Medeiros S., Gérard F., Veshchikov N., Lerman L., Markowitch O.* Breaking Kalyna 128/128 with Power Attacks, in Carlet, C., Hasan, M., Saraswat, V. (eds), *Security, Privacy, and Applied Cryptography Engineering. SPACE 2016. Lecture Notes in Computer Science,* Vol. 10076. Springer, Cham. Available at: https://doi.org/10.1007/978-3-319-49445-6_23.
13. *Jeon Y., Yoon J.W.* Filtering-Based Correlation Power Analysis (CPA) with Signal Envelopes Against Shuffling Methods, You, I. (eds), *Information Security Applications. WISA 2020. Lecture Notes in Computer Science,* Vol. 12583. Springer, Cham. Available at: https://doi.org/10.1007/978-3-030-65299-9_29.
14. *Lo O., Buchanan W.J., Carson D.* Power analysis attacks on the AES-128 S-box using differential power analysis (DPA) and correlation power analysis (CPA), *Journal of Cyber Security Technology,* 2016, 1 (2), pp. 88-107. Available at: https://doi.org/10.1080/23742917.2016.1231523.
15. *Xin J., Du Z.* Template attack based on uBlock cipher algorithm, *Frontiers in Computing and Intelligent Systems,* 2023, 3 (1), pp. 90-93. Available at: https://doi.org/10.54097/fcis.v3i1.6031.
16. GOST R 34.12-2015 Informatsionnaya tekhnologiya. Kriptograficheskaya zashchita informatsii. Blochnye shifry [GOST R 34.12-2015 Information technology. Cryptographic protection of information. Block ciphers]. Available at: URL: https://tc26.ru/standard/gost/GOST_R_3412-2015.pdf.

17. Statistical leakage simulator for the ARM M0 family ELMO. Available at:https://github.com/scaresearch/ELMO.
18. *Welch D*. Thumbulator. Available at: https://github.com/dwelch67/thumbulator.git.
19. CSA ISO/IEC 17825-2018 Information technology - Security techniques - Testing methods for the mitigation of non-invasive attack classes against cryptographic modules.
20. *Goodwill G., Jun B., Jaffe J. and Rohatgi P*. A testing methodology for side-channel resistance validation, *NIST Non-Invasive At-tack Testing Workshop,* 2011.
21. *Cooper J., DeMulder E., Goodwill G., Jaffe J., Kenworthy G. and Rohatgi P.* Test vector leakage assessment (tvla) methodology in practice, *International Cryptographic Module Conference,* 2013.

Статью рекомендовал к опубликованию д.т.н., профессор А.Н. Целых.

**Малявина Виктория Олеговна** – Южный федеральный университет; e-mail: malyavina@sfedu.ru; г. Таганрог, Россия; тел.: 88634371905; кафедра безопасности информационных технологий им. О.Б. Макаревича; выпускник.

**Маро Екатерина Александровна** – e-mail: eamaro@sfedu.ru; тел.: 88634371905; кафедра безопасности информационных технологий им. О.Б. Макаревича; к.т.н.; доцент.

**Malyavina Viktoriya Olegovna** – Southern Federal University; e-mail: malyavina@sfedu.ru; Taganrog, Russia; phone: +78634371905; the Department of Information Security; alumnus.

**Maro Ekaterina Aleksandrovna** – e-mail: eamaro@sfedu.ru; phone: +78634371905; the Department of Information Security; cand. of eng. sc.; associate professor.

УДК 004.272.44 DOI 10.18522/2311-3103-2024-5-173-185

**Д.А. Сорокин, А.В. Касаркин**

## ОБЗОР МОДЕЛЕЙ КОММУТАЦИОННЫХ ПОДСИСТЕМ ЦИФРОВЫХ ФОТОННЫХ ВЫЧИСЛИТЕЛЬНЫХ УСТРОЙСТВ

*Рассматриваются варианты организации подсистемы коммутации цифровых фотонных вычислительных устройств, основной задачей которой является обеспечение возможности организации эффективных вычислений при решении задач различных проблемных областей. По мнению авторов, цифровые фотонные вычислители должны обрабатывать информацию в структурной парадигме вычислений. Данная парадигма принципиально отличается от классической фон-Неймановской парадигмы, поскольку в ней передача данных между функциональными элементами не расторжима с обработкой. Поэтому проблематика построения подсистемы коммутации в разрабатываемых цифровых фотонных вычислительных устройствах – одна из ключевых. Данная подсистема должна обрабатывать информационные зависимости между выполняемыми операциями не только во времени, но и в пространстве. Только в этом случае обработка данных в фотонных вычислительных системах будет выполняться с производительностью, превосходящей на два и более десятичных порядка производительность самых современных электронных вычислительных систем. Рассматриваются вопросы обеспечения потокового обмена данными между функциональными устройствами в цифровом фотонном вычислителе. Авторы разработали и проанализировали в базисе фотонной логики модели коммутационных устройств и способы организации коммутационной подсистемы при выполнении последовательной обработки данных. В ходе исследований было установлено, что структурная организация вычислений в цифровых фотонных вычислителях возможна при обеспечении обмена данными посредством пространственной коммутации входных и выходных каналов функциональных устройств. При реализации цифровых фотонных вычислителей как универсальных устройств, ориентированных на широкий класс задач, наиболее удобными для организации вычислительных структур будут иерархический и иерархическо-кольцевой варианты подсистемы коммутации. Однако данные варианты характеризуются высокими накладными расходами на построение коммутаторов. Поэтому в проблемно-ориентированных фотонных вычислителях, предназначенных для решения сильносвязанных задач с высокой удельной производительностью, более предпочтительно применение ортогональной или тороидальной подсистемы коммутации. В этом случае должна обеспечиваться непосредственная пространственная коммутация между функциональными устройствами одной группы,*

*а также между группами. Данные варианты характеризуются более высокими требованиями к качеству формирования физических каналов между коммутаторами и функциональными устройствами, а также между самими коммутаторами.*

*Цифровое фотонное вычислительное устройство; коммутационные устройства ЦФВУ; структурная парадигма вычислений.*

**D.A. Sorokin, A.V. Kasarkin**

## OVERVIEW OF SWITCHING SUBSYSTEM MODELS FOR DIGITAL PHOTONIC COMPUTING DEVICES

*This article examines options for organizing the switching subsystem of digital photonic computing devices, whose main task is to enable efficient computations in various problem domains. According to the authors, digital photonic computers should process information within a structural computing paradigm. This paradigm fundamentally differs from the classical von Neumann paradigm, as data transfer between functional elements is inseparable from processing. Therefore, developing a switching subsystem in digital photonic computing devices is a critical challenge. This subsystem must handle data dependencies between operations not only in time but also in space. Only under these conditions can data processing in photonic computing systems achieve performance that exceeds the performance of the most advanced electronic computing systems by two or more decimal orders. The article addresses issues of streaming data exchange between functional devices in a digital photonic computer. The authors developed and analyzed switching device models and methods for organizing the switching subsystem for sequential data processing, using a basis of photonic logic. The research established that structural organization of computations in digital photonic computers is feasible when data exchange is achieved through spatial switching of input and output channels of functional devices. In implementing digital photonic computers as universal devices aimed at a wide range of tasks, hierarchical and hierarchical-ring variants of the switching subsystem organization are most suitable for forming computational structures. However, these variants are characterized by high overhead for constructing switches. Therefore, in problem-oriented photonic computers designed for solving highly interconnected tasks with high specific performance, the use of orthogonal or toroidal switching subsystems is preferred. In this case, direct spatial switching between functional devices within a group, as well as between groups, should be ensured. These variants have higher requirements for the quality of physical channels formed between switches and functional devices, as well as between the switches themselves.*

*Digital photonic computing device; switching devices of DPCD; structural computing paradigm.*

**Введение.** Современная микроэлектроника практически достигла технологических пределов своего развития, обусловленных физическими ограничениями размеров транзисторов [1]. После взрывного роста производительности вычислительных систем за счёт увеличения логической ёмкости микросхем и тактовых частот работы в 90-е годы прошлого века и в начале 2000-х в настоящее время мы наблюдаем только медленное развитие за счёт масштабирования количества обрабатывающих ядер. При этом архитектуры современных вычислительных систем в той или иной степени являются развитием фон-Неймановской концепции, что существенно снижает реальную производительность при решении трудоемких прикладных задач. Применение альтернативных способов организации обработки данных, таких как структурная парадигма [2] в системах, способно поднять реальную производительность до уровня 50-90% от пиковой, однако не решает проблему отсутствия возможности качественного наращивания вычислительной мощности, поскольку упирается в те же технологические барьеры микроэлектроники.

Преодоление подобных ограничений возможно только при создании принципиально новых вычислительных машин. Одним из таких перспективных вариантов являются цифровые фотонные системы [3–5], в которых вычисления выполняются с помощью светового потока, излучаемого лазером. Обработка данных фотонными логическими вентилями, такими как NOT, AND, OR и их производными, а также построенными на их базе триггерами и функциональными устройствами, способна обеспечить точность вычислений, аналогичную точности цифровой микроэлектроники. При этом работа на тактовых частотах терагерцового уровня позволит получить качественный рост производительности относительно самых современных вычислительных систем.

В настоящее время ведутся разработки как технологий построения элементной базы цифровых фотонных микросхем, так и технологий организации эффективных вычислений на их основе. В предыдущих работах [6, 7] был проанализирован вариант построения архитектуры цифровых фотонных вычислительных устройств (ЦФВУ), выполняющей обработку потоков операндов в структурной парадигме вычислений. Рассмотрены принципы обеспечения быстродействия и точности решения задач на ЦФВУ при выбранном способе представления данных. Предложены подходы к построению функциональных устройств в базисе фотонной логики, выполняющих арифметические операции в различных форматах представления данных. Однако в структурной парадигме необходимо соблюдать следующее правило: *процесс передачи данных между функциональными элементами реализуемых вычислительных структур не расторжим с процессом обработки*.

Для сравнения в многопроцессорных системах с традиционной архитектурой эти процессы разделены. Межпроцессорные обмены являются программной надстройкой, соединяющей множество независимо работающих процессоров, и зачастую для выполнения требуют выделенного времени при решении задач, что значительно снижает производительность систем. Это обусловлено тем, что множество независимых процессов обмена данными в традиционных вычислительных системах в общем случае конфликтно, поскольку не обеспечивается детерминизм вычислительных параллельных процессов [8].

Параллельная обработка данных на ЦФВУ возможна с соблюдением сформулированного выше правила, но требует аппаратной реализации множества информационных связей между работающими функциональными устройствами либо с помощью непосредственного соединения, либо посредством некоторого программно настраиваемого коммутационного ресурса.

Стоит отметить, что объединение функциональных элементов ЦФВУ в сложные вычислительные структуры порождает проблему обеспечения передачи данных на сверхвысоких тактовых частотах без искажений. Паллиативные подходы к организации безошибочной передачи в ЦФВУ, такие как контроль чётности или передача с квитанциями, нецелесообразно применять в структурной парадигме вычислений, поскольку требуют дополнительных аппаратных затрат и делают поток данных нерегулярным и разреженным. Поэтому необходимо разделение ЦФВУ на локальные участки, внутри которых будет обеспечена синхронная передача за счет прямой коммутации при соблюдении одинаковых длин трасс и низкого коэффициента затухания светового сигнала. Внутри таких локальных участков возможно использование коммутационного оборудования, настраиваемого на этапе программирования либо в процессе решения задачи при минимальных аппаратных затратах на синхронизацию данных.

В свою очередь между локальными участками ЦФВУ возможны длинные трассы передачи световых сигналов, что неминуемо приведет к рассинхронизации потоков данных. Поэтому между локальными участками необходимы устройства буферизации и синхронизации потоков данных, которые обеспечат синфазную подачу данных в ЦФВУ, а настройка коммутационного оборудования возможна только на этапе программирования.

В настоящей статье рассматриваются подходы к построению коммутационной подсистемы ЦФВУ, обеспечивающей перенаправление потоков данных между функциональными устройствами (ФУ) в вычислительных структурах и обладающей возможностью настройки информационных связей как в процессе программирования ЦФВУ, так и в процессе решения задач при организации ветвлений алгоритмов. Выбор варианта организации подсистемы коммутации ЦФВУ необходимо осуществлять через призму эффективности вычислительных структур, синтезируемых в базисе цифровой фотонной логики.

**Иерархическая коммутационная подсистема.** В работах [6, 7] рассматривались варианты реализации коммутационной подсистемы с иерархической топологией на основе универсальных полнодуплексных коммутаторов, мультиплексирующих и демультиплексирующих входные и выходные каналы данных по принципу «все со всеми», когда для каждого входа функционального устройства обеспечивается подключение к любому из выходов других функциональных устройств и наоборот. Это позволяет эффективно задействовать ресурс и реализовать вычислительную структуру, соответствующую ин-

формационному графу без редукционных преобразований [9], если число необходимых функциональных устройств больше или равно числу вершин информационного графа. Однако, как показывают исследования, в реальных прикладных задачах число вершин информационного графа, как правило, много больше числа доступных функциональных устройств не только в пределах одного ЦФВУ, но и в пределах высокопроизводительных систем, построенных на основе ЦФВУ. Для решения таких задач синтезируются вычислительные структуры, соответствующие редуцированному по производительности информационному графу [10], а обработка данных выполняется структурно-процедурным методом [11]. При этом информационные зависимости, не вошедшие в структурную компоненту, в общем случае реализуются через оперативную память. Поэтому аппаратная избыточность полнодуплексных коммутаторов неизбежно приведёт к падению удельной производительности.

Более эффективное использование аппаратного ресурса в иерархической коммутационной подсистеме при реализации различных задач, характеризующихся большим числом информационных вершин относительно необходимых ФУ в системе, возможно при применении полудуплексных коммутаторов, например как показано на рис. 1.

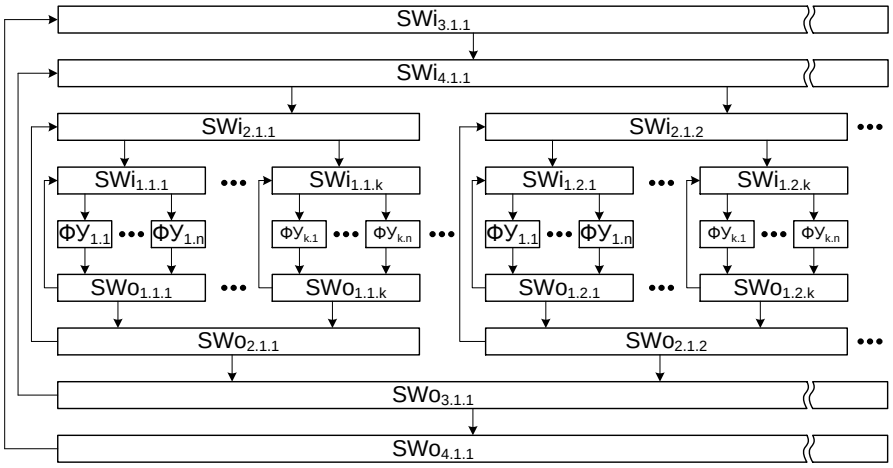



*Рис. 1. Вариант построения иерархической коммутационной подсистемы ЦФВУ с полудуплексными коммутаторами*

В приведённом варианте коммутационной подсистемы: $SWi_{i,j,k}$ – входные коммутаторы, $SWo_{i,j,k}$ – выходные коммутаторы, $ФУ_{k.n}$ – арифметико-логические функциональные устройства. Адресация коммутаторов определяется уровнем иерархии i, номером группы j, которая подключена к одному вышестоящему коммутатору, и номером коммутатора k в группе. Функциональные устройства разделены по группам, каждая из которых подключена к коммутаторам первого уровня (i=1). Коммутаторы являются устройствами типа «мультиплексор/демультиплексор»

Представленная коммутационная система включает в себя два типа коммутаторов. К первому типу относятся коммутаторы, которые могут выполнять в процессе решения задачи функции условных переходов, либо на этапе синтеза формировать вычислительную структуру. Ко второму типу относятся коммутаторы, предназначенные только для синтеза вычислительных структур в соответствии с информационным (исходным или редуцированным) графом задачи [10] на этапе программирования ЦФВУ. Для этого на их управляющие входы должны быть поданы конфигурационные параметры, сформированные на этапе трансляции программы ЦФВУ. Конфигурационные параметры не могут меняться в процессе решения задачи.

Рассмотрим фрагмент ЦФВУ с иерархической коммутационной системой, состоящей из трёх групп ФУ, обрабатывающих данные в формате 32-разрядной беззнаковой фиксированной запятой по три сумматора, умножителя и делителя, представленный на рис. 2.

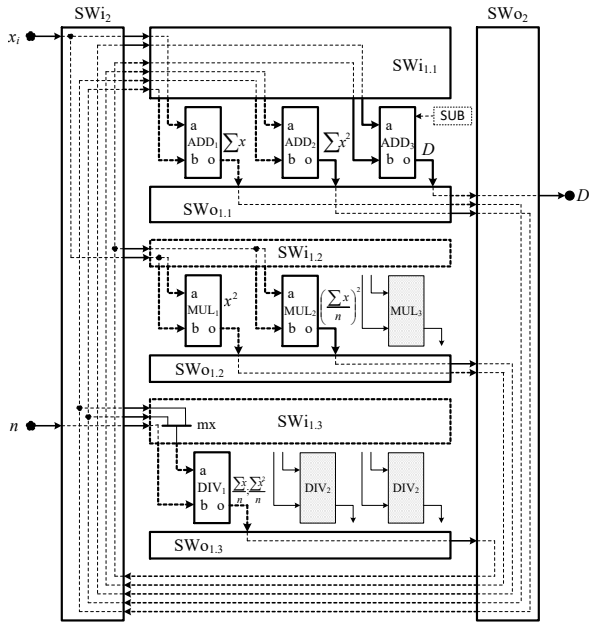



*Рис. 2. Вычислительная структура функции D с иерархической подсистемой коммутации*

На данном фрагменте показана конвейерная реализация вычислительной структуры функции расчёта дисперсии

$$D = \frac{\sum_{i=1}^{n} x_i^2}{n} - \left(\frac{\sum_{i=1}^{n} x_i}{n}\right)^2,$$

которая является оптимизированной версией структуры из [6].

Вычислительная структура расчёта *D* обрабатывает поток входных операндов $x_i$ за *f(i)* тактов. Число тактов *f(i)* определяется типом обработки данных (конвейер или процедура) и форматом представления данных (параллельный, последовательный и т.п.). При последовательной подаче данных старшими разрядами вперёд [12, 13] обработка в структуре *D* будет выполняться за $32 \cdot i$ такта. Однако, если выполнить два деления по очереди на одном делителе, функция *D* будет рассчитана за $32 \cdot (i+1)$ такта, то есть практически без потери в производительности, но будут уменьшены аппаратные затраты на один делитель. Это является существенным выигрышем, поскольку аппаратная реализация устройства деления в базисе ЦФВУ требует в 2,5 раза ресурса больше, чем устройство умножения, и в 15 раз больше, чем устройство суммирования.

Представленный на рисунке 2 фрагмент ЦФВУ содержит на первом уровне три входных коммутатора $SWi_{1.1}$-$SWi_{1.3}$ и три выходных коммутатора $SWo_{1.1}$-$SWo_{1.3}$, все первого типа. На втором уровне один входной коммутатор $SWi_2$ и один выходной $SWi_2$, оба второго типа. Массив ФУ разделён на три группы: сумматоры $ADD_1$-$ADD_3$, умножители $MUL_1$-$MUL_3$ и делители $DIV_1$-$DIV_3$. Неиспользуемые ФУ в вычислительной структуре *D* на рис. 2 заштрихованы. Для простоты изложения на рис. 2 не показаны элементы системы синхронизации, которые также необходимы для выполнения алгоритма.

Сформированные при программировании ЦФВУ в коммутаторах информационные каналы показаны пунктирными линиями. После того как все операнды $\boldsymbol{x_i}$ поступят на вход вычислительной структуры, на сумматорах $ADD_1$ и $ADD_2$ сформируются суммы $\sum x$ и $\sum x^2$ и через мультиплексор mx коммутатора $SWi_{1.3}$ будут последовательно поданы на вход **a** делителя $DIV_1$. На вход **b** $DIV_1$ подаётся переменная **n**. На выходе **o** делителя $DIV_1$ будут сформированы значения $\sum x/n$ и $\sum x^2/n$. Данные значения поступают на входы $ADD_3$, настроенного при программировании ЦФВУ на вычитание.

Здесь был показан лишь один из возможных вариантов формирования вычислительной структуры, и, в зависимости от возможностей и параметров коммутационной подсистемы ЦФВУ, вычислительная структура может значительно отличаться, например, возможно группировать ФУ не по их типу, а одну группу поместить весь набор различных ФУ (сумматоры, умножители и т.д.). Так как в пределах каждой группы ФУ устройства имеют большие возможности коммутации при решении некоторых классов задач, такой подход к группировке может быть более эффективным.

Несмотря на то что иерархическая организация подсистемы коммутации с применением полудуплексных коммутаторов требует существенно меньших аппаратных затрат по сравнению с универсальными полнодуплексными коммутаторами, в данном подходе имеется ряд принципиальных проблем при организации информационных обменов в вычислительных структурах решения различных сильносвязанных задач. Например, в большинстве случаев будет критичной организация передачи множество данных от одних ФУ к другим через несколько иерархических уровней. В рассмотренном примере видно, что при переходе на коммутатор более высокого уровня иерархии плотность коммутируемых зависимостей возрастает. Если для коммутаторов первого уровня количество выходных каналов определяется количеством входов одной группы ФУ, то для коммутаторов второго уровня количество выходных каналов определяется количеством входов всех групп ФУ, которые они объединяют. В результате при синтезе на ЦФВУ более сложных вычислительных структур, требующих применения десятков и сотен различных ФУ, коммутационная нагрузка на более старшие иерархические уровни будет значительно возрастать. Также могут возникать коллизии при выполнении вычислений, что потребует либо введения пауз в процесс обработки, либо применения полнодуплексных коммутаторов на втором и более высоких уровнях иерархии.

**Кольцевая коммутационная подсистема.** Многие вычислительно трудоёмкие задачи из области линейной алгебры, геофизики, криптографии и другие решаются итерационными алгоритмами обработки данных. Вычислительные структуры таких задач можно эффективно реализовывать с использованием кольцевой коммутационной подсистемы. Рассмотрим на рис. 3 пример типового фрагмента вычислительной структуры в ЦФВУ с кольцевой коммутацией. Подобная вычислительная структура может быть синтезирована для выполнения таких алгоритмов линейной алгебры, как прямой ход Гаусса или прямой ход LU-декомпозиции квадратной матрицы [14].

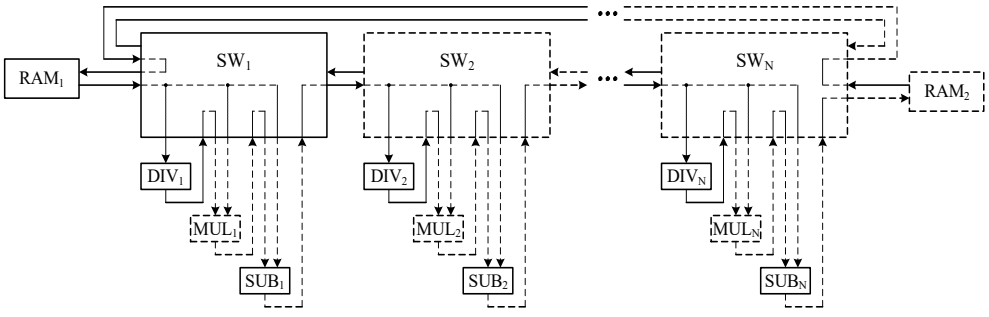



*Рис. 3. Типовой фрагмент вычислительной структуры задачи линейной алгебры*

В приведённом на рис. 3 варианте коммутационной подсистемы $SW_1$-$SW_N$ – коммутаторы. ФУ разделены по группам, состоящим из делителя $DIV_i$, умножителя $MUL_i$ и вычитателя $SUB_i$. В каждой группе все входы и выходы ФУ подключены к соответствующему коммутатору $SW_i$. Данная коммутационная система включает в себя только полнодуплексные коммутаторы, которые на этапе синтеза формируют вычислительную структуру, а в процессе решения задачи обрабатывают функции условных переходов и т.п. Каждый коммутатор $SW_i$ соединен только с 2 соседними коммутаторами $SW_{i-1}$ и $SW_{i+1}$.

Разложение матрицы на треугольные подматрицы является основой многих алгоритмов, связанных с численным решением систем линейных алгебраических уравнений. Если представленный фрагмент вычислительной структуры реализует прямой ход LU-разложения квадратной матрицы $A$ в виде произведения $LU=A$, то структурной парадигме вычислений конвейерная реализация предполагает, что $i$-я ступень конвейера соответствует одной итерации алгоритма и реализуется на ФУ $i$-го коммутатора. При программировании ЦФВУ в коммутаторах $SW_1$-$SW_N$ формируются информационные каналы (показаны пунктирными линиями). Для простоты изложения на рисунке не показаны элементы системы синхронизации, которые так же необходимы для выполнения алгоритма. Данные из памяти $RAM_1$ через систему сопряжения с внешней памятью [6] поступают на коммутатор $SW_1$, который подаёт их в обработку на ФУ первой группы. Результат работы группы ФУ отправляется через коммутатор $SW_1$ в коммутатор $SW_2$ и подаётся на обработку ФУ второй группы. В это же время в коммутатор $SW_1$ подгружаются новые данные из $RAM_1$. Так по цепочке данные непрерывным потоком перемещаются через коммутаторы от одной группы ФУ к следующей.

При количестве ступеней конвейера $N<M$, где M – размерность матрицы, матрица обрабатывается итерационно. Число итераций определяется как ]M/N[. На первой итерации исходная матрица читается из $RAM_1$, а промежуточные данные после прохождения по всем ступеням конвейера записываются в $RAM_2$. На второй итерации данные из $RAM_2$ после прохождения по вычислительному конвейеру записываются в $RAM_1$. Процесс повторяется, пока не будут выполнены все итерации LU-разложения.

Однако стоит отметить, что подсистема коммутации эффективна только для задач, которые характеризуются прямой информационной зависимостью между $i$ и $i+1$ ступенями вычислительной структуры. В то же время многие вычислительно трудоёмкие сильносвязанные задачи характеризуются более сложными информационными зависимостями, в общем случае определяемыми алгоритмом обработки. Рассмотрим такой вариант на примере вычислительно трудоемкого фрагмента задачи свёрточной нейронной сети [15].

Основными операциями свёрточной нейронной сети являются вычисления вида $\sum p \cdot w$. Напрямую реализовать сверточную нейронную сеть на ЦФВУ с кольцевой коммутационной подсистемой, как на рис. 3, затруднительно из-за того, что каждой группе ФУ нужен свой набор данных из оперативной памяти ЦФВУ. Для преодоления этой проблемы необходимы дополнительные коммутаторы, обеспечивающие сопряжение между подсистемой взаимодействия с внешней оперативной памятью и коммутаторами групп ФУ. При этом организация вычислительных структур в ЦФВУ, ориентированных на реализацию свёрточных нейросетей, будет возможна в случае совмещения в коммутационной подсистеме иерархического и кольцевого подхода [16]. Для ЦФВУ с такой коммутационной подсистемой на рис. 4 показан пример вычислительной структуры, состоящей из двух ядер нейронной сети.

Для реализации вычислительной структуры двух свёрточных ядер в иерархическо-кольцевой ЦФВУ необходимо задействовать 18 групп ФУ: $<MUL_{1.1};ADD_{1.1}>$ - $<MUL_{1.9};ADD_{1.9}>$ и $<MUL_{2.1};ADD_{2.1}>$ - $<MUL_{2.9};ADD_{2.9}>$, 18 групповых полнодуплексных коммутаторов $SW_{1.1.1}$-$SW_{1.1.9}$ и $SW_{1.2.1}$-$SW_{1.2.9}$, два коммутатора второго уровня $SW_{2.1.1}$ и $SW_{2.1.2}$, один коммутатор третьего уровня $SW_{3.1.1}$, также обеспечивающий сопряжение с подсистемой внешней оперативной памяти RAM.

Вычислительная структура выполняет свёртку окном 3x3 над матрицами пикселей размерностью $N$ по формуле

$$S_k^c = \sum_{i=1}^{9} w_i^c \cdot p_{i,k} \ ,$$

где $k$ – количество свёрток, для каждого $c$-го свёрточного ядра, определяемое как $N^2$;

$c$ – порядковый номер свёрточного ядра в нейроне;

$w^c{}_i$ – коэффициенты $c$-го ядра;

$p_{i,k}$ – значения пикселей изображения.

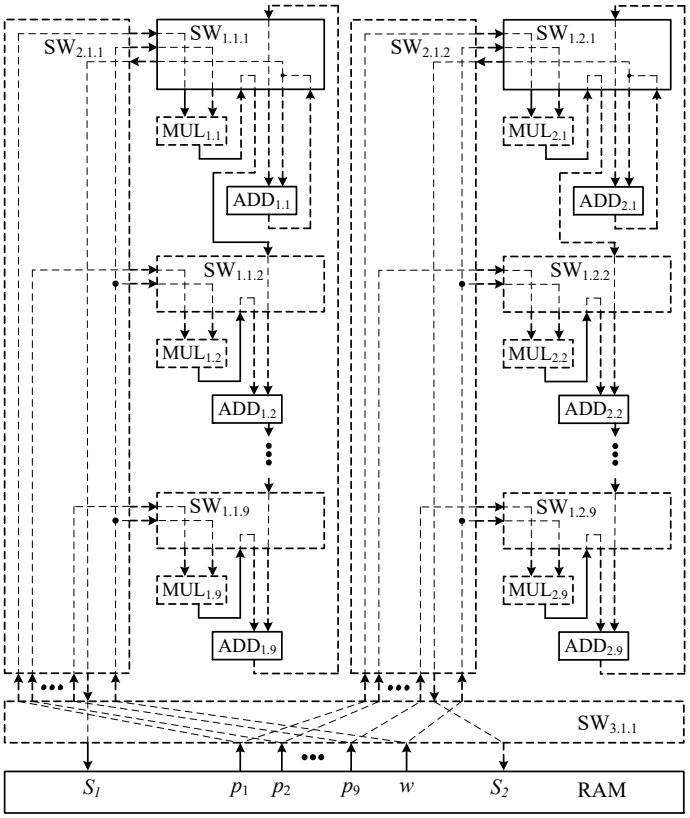



*Рис. 4. Вычислительная структура двух свёрточных ядер на иерархическо-кольцевой ЦФВУ*

Данная вычислительная структура характеризуется тем, что каждое ядро в соответствии с алгоритмом нейросети обрабатывает потоки данных $p_{i,k}$ своими девятью коэффициентами $w^c_i$. Стоит отметить, что в режиме исполнения нейросети коэффициенты $w^c_i$ не меняются, а в режиме обучения частота изменения данных коэффициентов пренебрежимо мала по сравнению с количеством выполняемых операций свёртки [17]. Поэтому их целесообразно загружать по единому каналу в регистры умножителей $MUL_{1.1}$- $MUL_{2.9}$, запоминая свой набор $w^c_i$ для каждого ядра.

Проведенные исследования показали, что для синтеза вычислительной структуры, приведённой на рис. 4, необходимый аппаратный ресурс на реализацию коммутаторов $SW_{2.1.1}$ и $SW_{2.1.2}$ будет сопоставим с суммарным аппаратным ресурсом коммутаторов $SW_{1.1.1}$-$SW_{1.1.9}$ и $SW_{1.2.1}$-$SW_{1.2.9}$. Дальнейшее масштабирование также приведёт к необходимости задействовать коммутаторы более высоких уровней иерархии. Поэтому неизбежен рост накладных расходов на организацию обмена данными. Дальнейшее масштабирование нейросетевых задач и сопоставимых им по связности задач из областей молекулярного моделирования, обработки графов, дистанционного зондирования Земли и многих других, на ЦФВУ с иерархическо-кольцевой коммутационной подсистемой приведёт к снижению удельной производительности на 10-15% на каждый дополнительный уровень иерархии.

**Ортогональная коммутационная подсистема.** Значительно более экономной в плане аппаратных затрат на коммутаторы по сравнению с иерархическо-кольцевой коммутационной подсистемой, а также менее подверженной проблемам коллизий за счет наличия большого числа пространственных связей между четырьмя соседними коммутаторами, является ортогональная коммутационная подсистема [18]. Также преимущества

ортогональной подсистемы ЦФВУ обеспечивают большое количество аппаратных связей между коммутаторами и, соответственно, большая вариативность возможного положения маршрутов, что позволяет балансировать нагрузку при обмене данными. Недостатком можно отметить неравномерность длин связей – чем ближе коммутатор к «краю» ЦФВУ, тем вероятней всего более длинные маршруты придется прокладывать, чтобы обеспечить соединение его группы ФУ с другими группами ФУ с соответствующими затратами на синхронизацию потоков данных.

Рассмотрим на рис. 5 вариант отображения вычислительной структуры двух свёрточных ядер на ЦФВУ с ортогональной коммутационной подсистемой.

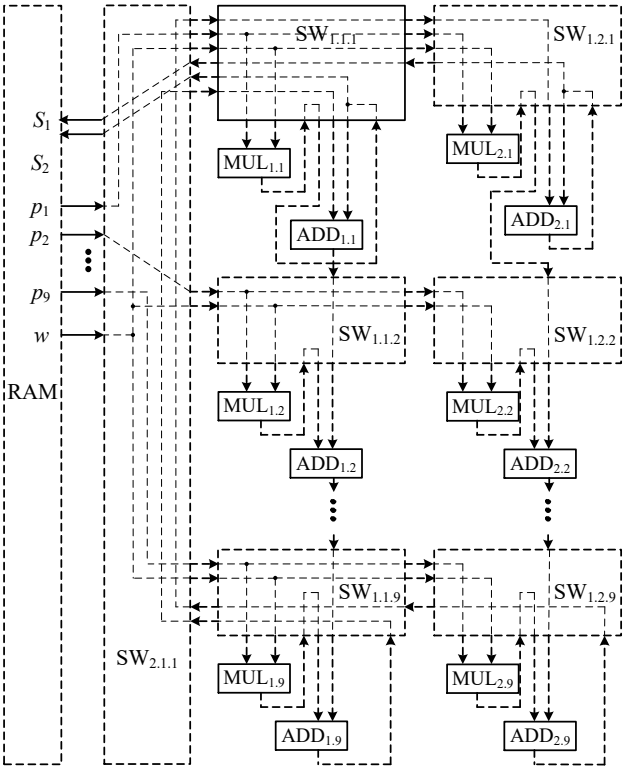



*Рис. 5. Вычислительная структура двух свёрточных ядер на ЦФВУ с ортогональной коммутационной подсистемой*

Для её реализации необходимо задействовать 18 групп ФУ: <$MUL_{1.1}$;$ADD_{1.1}$> - <$MUL_{1.9}$;$ADD_{1.9}$> и <$MUL_{2.1}$;$ADD_{2.1}$> - <$MUL_{2.9}$;$ADD_{2.9}$>, 18 групповых полнодуплексных коммутаторов $SW_{1.1.1}$-$SW_{1.1.9}$ и $SW_{1.2.1}$-$SW_{1.2.9}$ и один коммутатор второго уровня $SW_{2.1.1}$, также обеспечивающий сопряжение с подсистемой внешней оперативной памяти RAM.

Несмотря на то что в представленной вычислительной структуре коммутаторы $SW_{1.1.1}$-$SW_{1.1.9}$ и $SW_{1.2.1}$-$SW_{1.2.9}$ требуют примерно на 25% больше аппаратных затрат, чем каждый аналогичный коммутатор иерархическо-кольцевой коммутационной подсистемы, общие аппаратные затраты на ортогональную коммутацию значительно ниже, поскольку задействован только один коммутатор второго уровня иерархии.

Еще более снизить нагрузку на подсистему коммутации и обеспечить равномерность возможностей доступа всех коммутаторов можно, модернизировав подсистему ортогональной коммутации до подсистемы тороидальной коммутации [18]. Вычислительная структура для данного варианта ЦФВУ представлена на рис. 6.

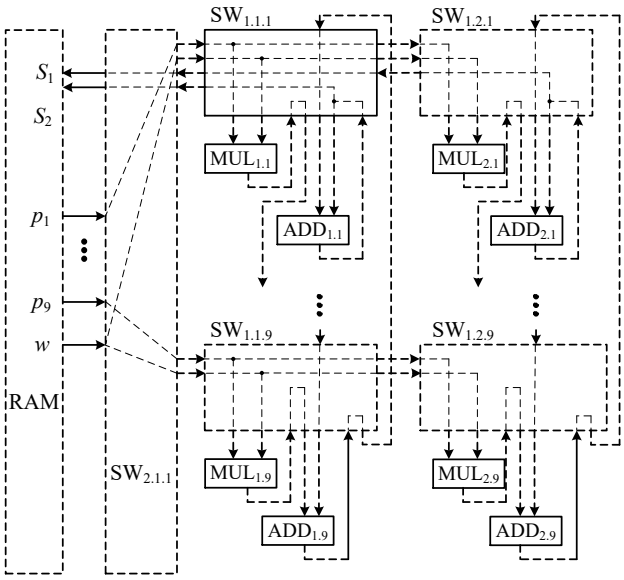



*Рис. 6. Вычислительная структура двух свёрточных ядер на ЦФВУ с тороидальной коммутационной подсистемой*

Сравнивая ЦФВУ с ортогональной и тороидальной коммутационной подсистемами, можно заметить, что при использовании тороидальной подсистемы коммутации связь между коммутаторами $SW_{1.1.1}$ и $SW_{1.1.9}$, $SW_{1.2.1}$ и $SW_{1.2.9}$ осуществляется напрямую, что также позволяет сократить аппаратные затраты на реализацию коммутатора $SW_{2.1.1}$ и значительно сокращает время прохождения сигналов. Подсистема тороидальной коммутации может быть эффективной за счёт обеспечения более равномерных условий прохождения сигналов между ФУ: здесь нет пограничных коммутаторов границ, для которых коммутаторам сложнее обеспечивать связь между ФУ. Однако такой вариант требует более качественного формирования кольцевых оптических каналов ЦФВУ, поскольку их длину будет определять геометрическое расстояние между коммутаторами $SW_{1.1.1}$ и $SW_{1.1.9}$, $SW_{1.2.1}$ и $SW_{1.2.9}$. Вероятнее всего потребуются обеспечение более высокой чистоты исполнения внутренней поверхности световодов и одинаковой их длины для обеспечения приемлемого коэффициента затухания [19,20]. В противном случае потребуется введение дополнительной синхронизации, например блоками выравнивания потоков операндов, описанными в [7].

**Заключение.** Проведенные исследования показали, что линейное масштабирование аппаратного ресурса ЦФВУ при решении сильносвязанных задач потенциально способно обеспечить близкий к линейному рост производительности. Для этого синтезируемые вычислительные структуры должны выполнять обработку данных в структурной парадигме. В зависимости от класса решаемых на ЦФВУ задач могут применяться различные подходы к организации коммутационной подсистемы обмена данными. Если разрабатываемые ЦФВУ будут ориентированы на задачи, в которых обрабатываются различного вида матрицы, более предпочтительным является кольцевой тип подсистемы коммутации, а для сложносоставных задач обработки нейронных сетей, молекулярного моделирования или дистанционного зондирования Земли более предпочтительными будут ортогональный или тороидальный типы коммутации.

*Исследование выполнено в рамках научной программы Национального центра физики и математики, направление №1 «Национальный центр исследования архитектур суперкомпьютеров. Этап 2023-2025».*

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Bérut Antoine.* Information and Thermodynamics: Experimental Verification of Landauer's Principle Linking Information and Thermodynamics. – URL: https://arxiv.org/pdf/1503.06537.pdf (дата обращения: 10.10.2024).
2. *Каляев А.В., Левин И.И.* Модульно-наращиваемые многопроцессорные системы со структурно-процедурной организацией вычислений. – М.: Янус-К, 2003. – 380 с.
3. *Степаненко С.А.* Фотонный компьютер: структура и алгоритмы, оценки параметров // Фотоника. – 2017. – № 7 / 67. – DOI: 10.22184/1993-7296.2017.67.7.72.83.
4. *Степаненко С.А.* Фотонная вычислительная машина. Принципы реализации. Оценки параметров // Доклады Академии наук. – 2017. – Т. 476, № 4. – С. 389-394. – DOI: 10.1134/S1064562417050234.
5. Chip war: China claims breakthrough in silicon photonics that could clear technical hurdle. – URL: https://www.scmp.com/tech/tech-war/article/3281156/chip-war-china-claims-breakthrough-silicon-photonics-could-clear-technical-hurdle (дата обращения: 10.10.2024).
6. *Сорокин Д.А., Левин И.И., Касаркин А.В.* Перспективная архитектура цифровой фотонной вычислительной машины // Известия ЮФУ. Технические науки. – 2022. – № 4 (2022). – С. 200-212. – DOI: 10.18522/2311-3103 2022.
7. *Sorokin D.A., Kasarkin A.V., Podoprigora A.V*. Elements of a Digital Photonic Computer // Supercomputing Frontiers and Innovations. – 2023. – Vol. 10, No. 2. – P. 62-76. – DOI: https://doi.org/10.14529/jsfi230205.
8. *Дейкстра Э.* Дисциплина программирования. – М.: Мир, 1978. – 278 с.
9. *Дордопуло А.И., Сорокин Д.А.* Методика сокращения аппаратных затрат в сложных системах при решении задач с существенно-переменной интенсивностью потоков данных // Известия ЮФУ. Технические науки. – 2012. – № 4 (129). – С. 194-199.
10. *Дордопуло А.И.* Применение методов редукции производительности для сокращения числа анализируемых вариантов параллельной программы // Вестник компьютерных и информационных технологий. – 2019. – № 9 (183). – С. 43-49. – DOI: 10.14489/vkit.2019.09.
11. *Kalyaev I.A., Levin I.I., Semernikov E.A., Shmoilov V.I.* Reconfigurable Multipipeline Computing Structures. – Published by Nova Science Publishers, Inc. (New York, USA), 2012. – 345 p. – ISBN 978-1-61942-854-6.
12. *Евстигнеев В.Г.* Недвоичные компьютерные арифметики // Электроника и информатика – 2005: Междунар. науч.-техн. конф. – М.: Ангсрем, 2006. – 774 с.
13. *B Devika Rani, Dr S Govindarajulu.* Implementation of modified vedic multiplier using on quaternary signed digit number system // Journal of Engineering Sciences. – 2022. – Vol. 13, Issue 12. – ISSN 0377-9254.
14. *Левин И.И., Пелипец А.В., Сорокин Д.А.* Решение задачи LU–декомпозиции на реконфигурируемых вычислительных системах: оценка и перспективы // Известия ЮФУ. Технические науки. – 2015. – № 6 (168). – С. 62-70.
15. *Николенко С., Кадурин А., Архангельская Е.* Глубокое обучение. – СПб.: Питер, 2018. – 480 с. – (Серия «Библиотека программиста»). – ISBN 978-5-496-02536-2.
16. *Дерюгин А.А.* Коммутаторы вычислительных систем // Вычислительные сети. Теория и практика. – 2006. – № 1 (8). – URL: https://network-journal.mpei.ac.ru/cgi-bin/main.pl?l=ru&n=8&pa=3&ar=1 (дата обращения: 10.10.2024).
17. *Шалев-Шварц Ш., Бен-Давид Ш.* Идеи машинного обучения: от теории к алгоритмам: пер. с англ. А.А. Слинкина. – М.: ДМК Пресс, 2019. – 436 с. – ISBN 978-5-97060-673-5.
18. *Хорошевский В.Г.* Архитектура вычислительных систем: учеб. пособие. – 2-е изд., перераб. и доп. – М.: Изд-во МГТУ им. Н.Э. Баумана, 2008. – 520 с. (Информатика в техническом университете). – ISBN 978-5-7038-3175-5.
19. *Коробейников А.Г., Гатчин Ю.А., Дукельский К.В., Тер-Нерсесянц Е.В.* Технологические методы снижения уровня оптических потерь в микроструктурированных волоконных световодах // Научно-технический вестник информационных технологий, механики и оптики. – 2014. – № 1 (89).
20. *Pavelyev V., Krivosheeva Y., Golovashkin D.* Genetic Optimization of the Y-Shaped Photonic Crystal NOT Logic Gate // Photonics. – 2023. – Vol. 10. – DOI: 10.3390/photonics10101173.

REFERENCES

1. *Bérut Antoine.* Information and Thermodynamics: Experimental Verification of Landauer's Principle Linking Information and Thermodynamics. Available at: https://arxiv.org/pdf/1503.06537.pdf (accessed 10 October 2024).
2. *Kalyaev A.V., Levin I.I.* Modul'no-narashchivaemye mnogoprotsessornye sistemy so strukturno-protsedurnoy organizatsiey vychisleniy [Modular-scalable multiprocessor systems with structural-procedural organization of computations]. Moscow: Yanus-K, 2003, 380 p.

3. *Stepanenko S.A.* Fotonnyy komp'yuter: struktura i algoritmy, otsenki parametrov [Photonic computer: structure and algorithms, parameter estimates], *Fotonika* [Photonics], 2017, No. 7 / 67. DOI: 10.22184/1993-7296.2017.67.7.72.83.
4. *Stepanenko S.A.* Fotonnaya vychislitel'naya mashina. Printsipy realizatsii. Otsenki parametrov [Photonic computing machine. Implementation principles. Parameter estimates], *Doklady Akademii nauk* [Reports of the Academy of Sciences], 2017, Vol. 476, № 4. – S. 389-394. – DOI: 10.1134/S1064562417050234.
5. Chip war: China claims breakthrough in silicon photonics that could clear technical hurdle. Available at: https://www.scmp.com/tech/tech-war/article/3281156/chip-war-china-claims-breakthrough-silicon-photonics-could-clear-technical-hurdle (accessed 10 October 2024).
6. *Sorokin D.A., Levin I.I., Kasarkin A.V.* Perspektivnaya arkhitektura tsifrovoy fotonnoy vychislitel'noy mashiny [Promising architecture of a digital photonic computing machine], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2022, No. 4 (2022), pp. 200-212. DOI: 10.18522/2311-3103 2022.
7. *Sorokin D.A., Kasarkin A.V., Podoprigora A.V.* Elements of a Digital Photonic Computer, *Supercomputing Frontiers and Innovations,* 2023, Vol. 10, No. 2, pp. 62-76. DOI: https://doi.org/10.14529/jsfi230205.
8. *Deykstra E.* Distsiplina programmirovaniya [Programming discipline]. Moscow: Mir, 1978, 278 p.
9. Dordopulo A.I., Sorokin D.A. Metodika sokrashcheniya apparatnykh zatrat v slozhnykh sistemakh pri reshenii zadach s sushchestvenno-peremennoy intensivnost'yu potokov dannykh [Methodology for reducing hardware costs in complex systems when solving problems with significantly variable data flow intensity], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2012, No. 4 (129), pp. 194-199.
10. *Dordopulo A.I.* Primenenie metodov reduktsii proizvoditel'nosti dlya sokrashcheniya chisla analiziruemykh variantov parallel'noy programmy [Application of performance reduction methods to reduce the number of analyzed variants of a parallel program], *Vestnik komp'yuternykh i informatsionnykh tekhnologiy* [Bulletin of Computer and Information Technologies], 2019, No. 9 (183), pp. 43-49. DOI: 10.14489/vkit.2019.09.
11. *Kalyaev I.A., Levin I.I., Semernikov E.A., Shmoilov V.I.* Reconfigurable Multipipeline Computing Structures. Published by Nova Science Publishers, Inc. (New York, USA), 2012, 345 p. ISBN 978-1-61942-854-6.
12. *Evstigneev V.G.* Nedvoichnye komp'yuternye arifmetiki [Non-binary computer arithmetic], *Elektronika i informatika – 2005: Mezhdunar. nauch.-tekhn. konf.* [Electronics and Information Science – 2005: International Scientific and Technical Conference]. Moscow: Angsrem, 2006, 774 p.
13. *B Devika Rani, Dr S Govindarajulu.* Implementation of modified vedic multiplier using on quaternary signed digit number system, *Journal of Engineering Sciences,* 2022, Vol. 13, Issue 12. ISSN 0377-9254.
14. *Levin I.I., Pelipets A.V., Sorokin D.A.* Reshenie zadachi LU–dekompozitsii na rekonfiguriruemykh vychislitel'nykh sistemakh: otsenka i perspektivy [Solution of the LU-decomposition problem on reconfigurable computing systems: assessment and prospects], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2015, No. 6 (168), pp. 62-70.
15. *Nikolenko S., Kadurin A., Arkhangel'skaya E.* Glubokoe obuchenie [Deep learning]. Saint Petersburg: Piter, 2018, 480 p. (Seriya «Biblioteka programmista» [Series "Programmer's Library"]). ISBN 978-5-496-02536-2.
16. *Deryugin A.A.* Kommutatory vychislitel'nykh sistem [Switches of computing systems], *Vychislitel'nye seti. Teoriya i praktika* [Computing networks. Theory and practice], 2006, No. 1 (8). Available at: https://network-journal.mpei.ac.ru/cgi-bin/main.pl?l=ru&n=8&pa=3&ar=1 (accessed 10 October 2024).
17. *Shalev-Shvarts Sh., Ben-David Sh.* Idei mashinnogo obucheniya: ot teorii k algoritmam [Machine learning ideas: from theory to algorithms]: trans. from engl. by. A.A. Slinkina. Moscow: DMK Press, 2019, 436 p. ISBN 978-5-97060-673-5.
18. *Khoroshevskiy B.G.* Arkhitektura vychislitel'nykh sistem: ucheb. posobie [Architecture of computing systems: a tutorial]. 2nd ed. Moscow: Izd-vo MGTU im. H.E. Baumana, 2008, 520 p. (Informatika v tekhnicheskom universitete [Informatics at a technical university]). ISBN 978-5-7038-3175-5.
19. *Korobeynikov A.G., Gatchin Yu.A., Dukel'skiy K.V., Ter-Nersesyants E.V.* Tekhnologicheskie metody snizheniya urovnya opticheskikh poter' v mikrostrukturirovannykh volokonnykh svetovodakh [Technological methods for reducing the level of optical losses in microstructured fiber optics], *Nauchno-tekhnicheskiy vestnik informatsionnykh tekhnologiy, mekhaniki i optiki* [Scientific and technical bulletin of information technologies, mechanics and optics], 2014, No. 1 (89).
20. *Pavelyev V., Krivosheeva Y., Golovashkin D.* Genetic Optimization of the Y-Shaped Photonic Crystal NOT Logic Gate, *Photonics,* 2023, Vol. 10. DOI: 10.3390/photonics10101173.

Статью рекомендовал к опубликованию д.т.н. Э.В. Мельник.

**Сорокин Дмитрий Анатольевич** – НИЦ супер-ЭВМ и нейрокомпьютеров; e-mail: jotun@inbox.ru; г. Таганрог, Россия; тел.: +79508668253; начальник отдела; к.т.н.

**Касаркин Алексей Викторович** – e-mail: kav589@mail.ru; тел.: +79045065636; научный сотрудник; к.т.н.

**Sorokin Dmitriy Anatolyevich** – Supercomputers and Neurocomputers Research Center; e-mail: jotun@inbox.ru; Taganrog, Russia; phone: +79508668253; chief of department; cand. of eng. sc.

**Kasarkin Alexey Viktorovich** – e-mail: kav589@mail.ru; phone: +79045065636; research scientist; cand. of eng. sc.

УДК 537.876 DOI 10.18522/2311-3103-2024-5-185-194

**М. Пленингер, С.В. Балакирев, М.С. Солодовник**

## МОДЕЛИРОВАНИЕ РАСПРЕДЕЛЕНИЯ НАПРЯЖЕННОСТИ ЭЛЕКТРИЧЕСКОГО ПОЛЯ В ПОЛНОСТЬЮ ОПТИЧЕСКОМ ЛОГИЧЕСКОМ КОМПАРАТОРЕ НА ОСНОВЕ ФОТОННОГО КРИСТАЛЛА GaAs

*Фотонные кристаллы – полупроводниковые структуры с фотонной запрещенной зоной – вызывают большой интерес у научного сообщества. Они представляют собой новый класс оптических материалов, обладающих пространственной периодической модуляцией диэлектрической проницаемости с периодом, близким к длине волны излучения. Интерес к этим структурам объясняется их значимостью для фундаментальных исследований взаимодействия излучения с веществом и потенциалом создания оптоэлектронных устройств следующего поколения. В данной работе представлены результаты моделирования компактного оптического логического компаратора на фотонном кристалле GaAs, работающем во втором окне прозрачности оптического волокна (длина волны 1.3 мкм). Модельный компаратор представляет собой среду с двумя входными и двумя выходными оптическими каналами. При вводе излучения в один из входов компаратора соответствующий выходной канал пропускает излучение, символизируя логическую единицу. В случае отсутствия сигналов на входных каналах либо ввода сигналов в оба входа, оба выходных канала не пропускают излучение, символизируя логические нули. Каналы в компараторе создаются с помощью пересекающихся волноводов, сформированных в двумерном фотонном кристалле GaAs, который состоит из набора цилиндрических кристаллов (столбцов) GaAs с диаметром от 130 до 170 нм, встроенных в вакуумную среду с периодом от 450 до 750 нм. Для обеспечения затухания электромагнитных волн, вводимых в компаратор в оба входных канала, в месте пересечения волноводов встроены дефектные столбцы GaAs с меньшим диаметром. Проведено исследование влияния диаметра столбцов и периода между столбцами фотонного кристалла GaAs на закономерности распространения электромагнитного излучения в среде оптического компаратора. На основании анализа отношения уровней интенсивности сигналов на входах и выходах устройства, установлено, что оптимальный диаметр столбцов GaAs и расстояние между ними, при которых структура в наибольшей степени соответствует требованиям работы оптического логического компаратора, составляет 155 и 600 нм соответственно.*

*Фотонный кристалл GaAs; оптический компаратор; моделирование.*

**M. Pleninger, S.V. Balakirev, M.S. Solodovnik**

## SIMULATION OF THE ELECTRIC FIELD STRENGTH DISTRIBUTION IN AN ALL-OPTICAL LOGIC COMPARATOR BASED ON THE GaAs PHOTONIC CRYSTAL

*Photonic crystals, semiconductor structures with a photonic band gap, are of great interest to the scientific community. They represent a new class of optical materials with spatial periodic modulation of permittivity with a period close to the wavelength of radiation. Interest in these structures is explained by their importance for fundamental studies of the interaction of radiation with matter and the potential for creating next-generation optoelectronic devices. This paper presents the results of modeling a compact optical logic comparator based on a GaAs photonic crystal operating in the second transparency window of an optical fiber (wavelength of 1.3 μm). The model comparator is a medium with two input and two*

*output optical channels. When radiation is input to one of the comparator inputs, the corresponding output channel transmits radiation, indicating a logical one. In the absence of signals on the input channels or when signals are input to both input channels, both output channels do not transmit radiation, indicating logical zeros. The channels in the comparator are created using intersecting waveguides formed in a two-dimensional GaAs photonic crystal, which consists of a set of cylindrical GaAs crystals (pillars) with a diameter of 130 to 170 nm, embedded in a vacuum medium with a period of 450 to 750 nm. To ensure attenuation of electromagnetic waves introduced into the comparator in both input channels, defective GaAs pillars with a smaller diameter are embedded at the intersection of the waveguides. The influence of the diameter and period between the GaAs photonic crystal pillars on the propagation patterns of electromagnetic radiation in the optical comparator medium is studied. Based on the analysis of the ratio of signal intensity levels at the inputs and outputs of the device, it is established that the optimal diameter of the GaAs pillars and the distance between them, at which the structure best meets the requirements of the logic comparator, is 155 and 600 nm, respectively.*

*GaAs photonic crystal; optical comparator; simulation.*

**Введение.** Фотонные кристаллы представляют собой современный тип неоднородных оптических материалов, отличительной особенностью которых является пространственная периодическая модуляция диэлектрической проницаемости. Период этой модуляции сопоставим с длиной волны излучения, а в спектре собственных электромагнитных состояний кристалла присутствует фотонная запрещенная зона [1]. Эта зона представляет собой диапазон частот, в котором излучение, распространяющееся в определенных направлениях фотонного кристалла (в зависимости от его размерности), экспоненциально теряет свою интенсивность. Следовательно, излучение, которое попадает на фотонный кристалл, будет значительно отражаться [2].

Внедрение фотонных кристаллов в качестве основы для создания различных оптических схем произвело революцию в оптической промышленности. От интегрированной фотоники до зондирования, от квантовой обработки информации до сбора солнечной энергии [3] – фотонные кристаллы обеспечивают беспрецедентный контроль над излучением и прокладывают путь для инновационных приложений и устройств [4]. Фотонные кристаллы могут иметь многочисленные практические и теоретические применения [5–9]. Они широко применяются в оптической интегральной схемотехнике, оптоэлектронных модуляторах, лазерах [10, 11] и биофотонике [12]. Важным направлением исследований фотонных кристаллов является разработка логических элементов фотонных интегральных схем на их основе. В частности, недавно были разработаны логические элементы «ИЛИ-НЕ» и «И-НЕ» на основе фотонных кристаллов [13, 14]. Также представлены результаты моделирования логического компаратора на основе фотонного кристалла арсенида галлия (GaAs) [15]. Однако до сих пор недостаточно исследований, посвященных фотонным кристаллам, работающим на практически значимой длине волны 1.3 мкм.

В данной работе проводится моделирование распределения напряженности электрического поля ($E$) в полностью оптическом логическом компараторе, который основан на фотонном кристалле с использованием столбцов полупроводникового кристалла GaAs и работает на длине волны 1.3 мкм. Выбор данной длины волны обусловлен тем, что она находится в диапазоне второго окна прозрачности оптического волокна и отличается нулевой дисперсией [16]. Кроме того, на этой длине волны излучают квантовые точки InGaAs, которые изготавливаются с помощью отлаженной и воспроизводимой эпитаксиальной технологии [17, 18].

**Описание модели оптического компаратора на основе фотонного кристалла GaAs.** Моделирование проводилось в программной среде COMSOL Multiphysics 6.1 с использованием модуля Wave Optics. С помощью данного модуля возможно изучение поперечных электрических (TE) волн, распространяющихся через фотонный кристалл. Для их моделирования используется скалярное уравнение для поперечной составляющей электрического поля $E_z$.

$$- \cdot \quad E_z - n^2 {k_0}^2 E_z = 0,$$

где $n$ – показатель преломления, а $k_0$ – волновое число в пространстве.

На основе фотонного кристалла возможно создание полностью оптического логического компаратора. Отличие оптического компаратора от компаратора, используемого в электронных микросхемах, состоит в том, что оптический компаратор имеет два входных и два выходных канала, тогда как электронный компаратор имеет три выходных канала. Для лучшего понимания работы оптического компаратора приведена таблица истинности данного логического элемента (табл. 1), где $I_1$ – первый входной канал, $I_2$ – второй входной канал; $O_1$ – первый выходной канал, $O_2$ – второй выходной канал.

Таблица 1

**Таблица истинности оптического логического компаратора**

| $I_1$ | $I_2$ | $O_1$ | $O_2$ | Статус |
|---|---|---|---|---|
| 0 | 0 | 0 | 0 | $I_1 = I_2$ |
| 0 | 1 | 0 | 1 | $I_1 < I_2$ |
| 1 | 0 | 1 | 0 | $I_1 > I_2$ |
| 1 | 1 | 0 | 0 | $I_1 = I_2$ |

В качестве моделируемого геометрического объекта был задан фотонный кристалл, состоящий из сонаправленных цилиндрических столбцов GaAs, расположенных в вакуумной среде и составляющих структуру гексагональной решетки (рис. 1). Известно, что фотонные кристаллы со структурными элементами в форме гексагонов обладают особым интересом в связи с возможностью максимального отражения излучения с частотой, принадлежащей фотонной запрещенной зоне [19].

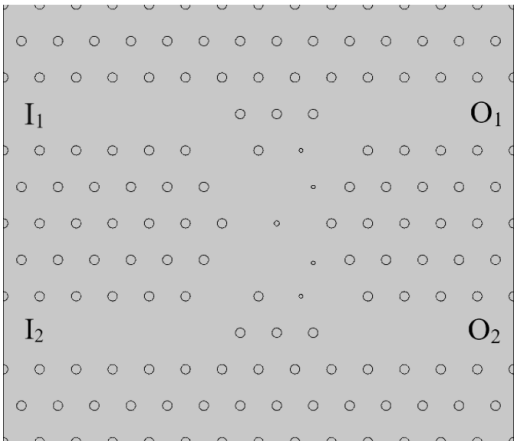



*Рис. 1. Топология моделируемого фотонного кристалла GaAs, на основе которого реализован оптический логический компаратор. $I_1$, $I_2$ – входы, $O_1$, $O_2$ – выходы компаратора. Кружками большего диаметра обозначены стандартные столбцы GaAs, кружками меньшего диаметра – дефектные столбцы, обеспечивающие взаимное гашение электромагнитных волн, поступающих на оба входа одновременно*

Диаметр столбцов GaAs ($d$) варьировался от 130 до 170 нм. Период фотонного кристалла ($a$), влияющий на глубину проникновения электромагнитного излучения, варьировался в диапазоне 450–750 нм. В модель также включены столбцы с меньшим диаметром в области пересечения волноводов для обеспечения достаточного затухания сигнала при двух открытых каналах. Такой подход позволяет усовершенствовать характеристики компаратора и расширить диапазон частот, на которых фотонный кристалл способен отражать излучение в области пересечения волноводов при обоих открытых каналах. Диаметр основных столбцов GaAs, при котором варьировался их период, составлял 155 нм. Диаметр меньших столбцов во всех случаях составлял 62 нм (4 столбца) и 83 нм (1 столбец) соответственно. Период структуры, при котором варьировался диаметр столбцов, составлял 600 нм.

Структура состоит из 15 столбцов в горизонтальном направлении и 13 столбцов в вертикальном направлении. Как показано на рис. 1, в фотонном кристалле, посредством удаления ряда столбцов, искусственно созданы два взаимопересекающихся волновода. Таким образом, оптический компаратор имеет два входных и два выходных канала. При подаче излучения в оба входных канала, излучение, распространяющееся по ним к области их пересечения, практически полностью затухает. Если излучение подается только в один из каналов (либо в первый, либо во второй), излучение проходит по волноводу с минимальными потерями [20].

**Результаты и обсуждение.** Проведено исследование распределения напряженности электромагнитного поля в оптическом компараторе при различных геометрических параметрах его структуры, таких как диаметр столбцов GaAs и период, т.е. расстояние между ними. В табл. 2 приведены значения напряженности электрического поля при различных диаметрах столбцов GaAs и при постоянном периоде структуры, равном 600 нм. Значения приведены при одном (один ввод) и двух (два ввода) открытых каналах.

Таблица 2

**Значения напряженности электрического поля (В/м) при различных диаметрах столбцов GaAs и периоде 600 нм**

| | Напряженность электрического поля (В/м) | | | |
|---|---|---|---|---|
| | Один ввод | | Два ввода | |
| Диаметр (нм) | На входе | На выходе | На входе | На выходе |
| 135 | 1.884 | 0.980 | 1.946 | 0.798 |
| 140 | 1.934 | 0.956 | 2.045 | 0.715 |
| 145 | 1.986 | 0.918 | 2.147 | 0.624 |
| 150 | 2.025 | 0.877 | 2.233 | 0.533 |
| 155 | 2.064 | 0.833 | 2.314 | 0.444 |
| 160 | 2.103 | 0.786 | 2.392 | 0.366 |
| 170 | 2.189 | 0.698 | 2.503 | 0.228 |
| 180 | 2.252 | 0.622 | 2.596 | 0.162 |
| 190 | 2.331 | 0.567 | 2.605 | 0.121 |

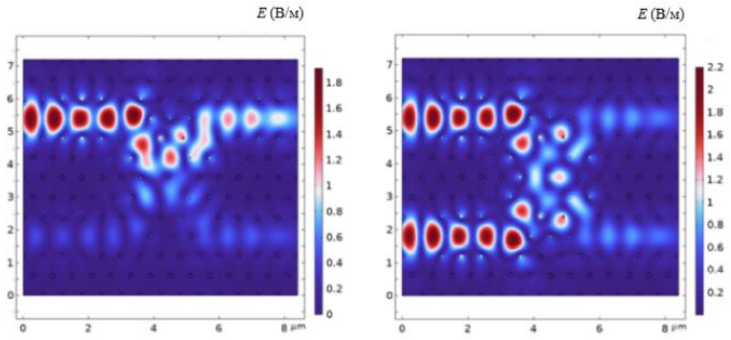
Аналогичная таблица построена для значений зависимости напряженности электрического поля при различных периодах структуры и неизменном диаметре столбцов GaAs, равном 155 нм (табл. 3).

Таблица 3

**Значения напряженности электрического поля (В/м) при различных расстояниях между столбцами GaAs и диаметре 155 нм**

| | Напряженность электрического поля (В/м) | | | |
|---|---|---|---|---|
| | Один ввод | | Два ввода | |
| Период (нм) | На входе | На выходе | На входе | На выходе |
| 450 | 1.931 | 0.022 | 2.099 | 0.004 |
| 500 | 2.654 | 0.232 | 3.141 | 0.299 |
| 550 | 2.110 | 1.541 | 1.858 | 1.681 |
| 600 | 2.064 | 0.833 | 2.314 | 0.444 |
| 650 | 2.051 | 0.064 | 2.017 | 0.050 |
| 700 | 2.209 | 0.137 | 2.202 | 0.218 |
| 750 | 1.990 | 0.271 | 1.962 | 0.495 |

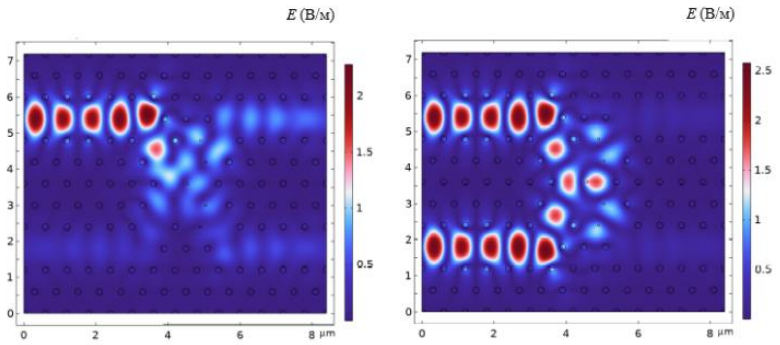
На рис. 2 представлены результаты моделирования оптического компаратора с диаметром столбцов фотонного кристалла, равным 140 нм, и периодом, равным 600 нм. Видно, что при вводе излучения в один канал излучение проходит по волноводу с определенными потерями (рис. 2,а) – напряженность электрического поля падает примерно в 2 раза (табл. 2). Несмотря на то, что при вводе излучения в один канал, в идеальном случае, напряженность поля на выходе должна сохранять такое же значение, как и на входе, добиться этого в случае данного оптического компаратора практически невозможно из-за неизбежных потерь. В связи с этим важно определить параметры фотонного кристалла, при котором затухание излучения при вводе в один канал будет минимальным, а при вводе в оба канала – максимальным. Так, при диаметре столбцов 140 нм и периоде между ними 600 нм излучение при вводе в оба канала затухает почти в три раза (рис. 2,б): напряженность электрического поля падает с 2.0 до 0.7 В/м (табл. 2).




а б

*Рис. 2. Полученное в результате проведенного моделирования распределение напряженности электрического поля в фотонном кристалле: а – при одном открытом канале, б – при обоих открытых каналах; d = 140 нм, a = 600 нм*

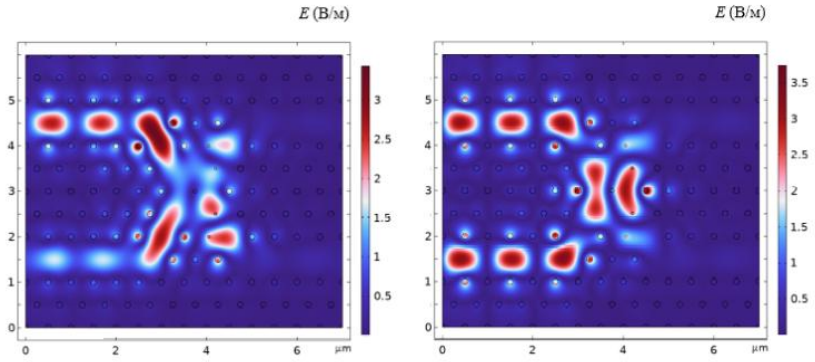
На рис. 3 представлены результаты моделирования оптического компаратора с диаметром столбцов, равным 170 нм, и периодом, равным 600 нм. Видно, что при двух открытых каналах излучение практически полностью затухает в месте пересечения волноводов (рис. 3,б) в соответствии с требуемыми критериями функционирования оптического логического компаратора: напряженность электрического поля на выходе падает более, чем на порядок, относительно значения на входе (табл. 2). Однако при одном открытом канале излучение доходит до выходного канала также с большими потерями (рис. 3,а): напряженность электрического поля падает с 2.2 до 0.7 В/м, т.е. примерно в 3 раза (табл. 2).




а б

*Рис. 3. Полученное в результате проведенного моделирования распределение напряженности электрического поля в фотонном кристалле: а – при одном открытом канале, б – при обоих открытых каналах; диаметр d = 170 нм, период a = 600 нм*

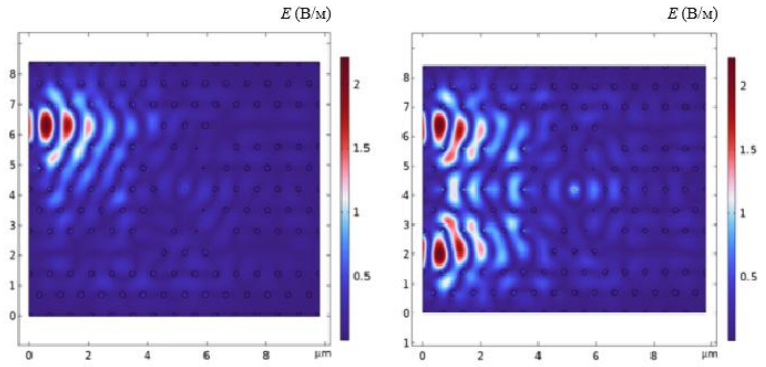
На рис. 4 представлены результаты моделирования оптического компаратора с диаметром столбцов, равным 155 нм, и периодом, равным 500 нм. Видно, что структура не удовлетворяет требованиям работы оптического компаратора, так как при одном открытом канале излучение не проходит по волноводу, а отражается от столбцов GaAs уменьшенного диаметра, расположенных в области пересечения волноводов (рис. 4,а); при этом напряженность электрического поля падает с 2.7 до 0.2 В/м (табл. 3).




а б

*Рис. 4. Полученное в результате проведенного моделирования распределение напряженности электрического поля в фотонном кристалле: а – при одном открытом канале, б – при обоих открытых каналах; d = 155 нм, a = 500 нм*

На изображениях с результатами моделирования, представленных на рис. 5, расстояние между столбцами GaAs составляет 700 нм при диаметре 155 нм. Снова можно сделать вывод, что такая конфигурация фотонного кристалла является неприемлемой, так как не позволяет реализовать оптический компаратор: излучение распространяется через всю структуру, а не только вдоль волноводов, и практически полностью рассеивается, не доходя до выходных каналов. Это связано с тем, что расстояние между столбцами в данном случае слишком велико для обеспечения достаточного отражения излучения с длиной волны 1.3 мкм.

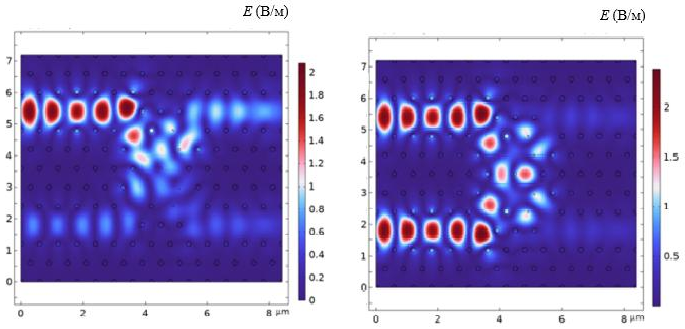



а б

*Рис. 5. Полученное в результате проведенного моделирования распределение напряженности электрического поля в фотонном кристалле: а – при одном открытом канале, б – при обоих открытых каналах; d = 155 нм, a = 700 нм*

На рис. 6 представлены результаты моделирования для оптимизированной структуры фотонного кристалла с диаметром столбцов 155 нм и периодом 600 нм. При одном открытом канале (рис. 6,а) излучение проходит через волновод с относительно малыми

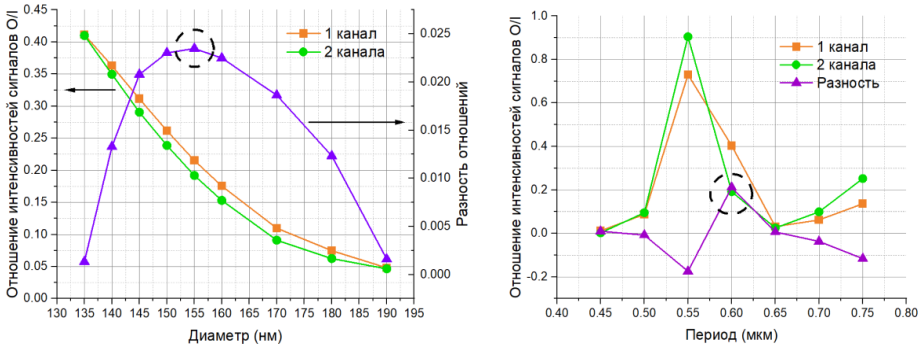
потерями (напряженность электрического поля падает с 2.1 до 0.8 В/м), но при этом при вводе излучения в оба канала (рис. 6,б) оно практически полностью затухает в области пересечения волноводов, а напряженность электрического поля падает с 2.3 до 0.4 В/м (табл. 3).




а б

*Рис. 6. Полученное в результате проведенного моделирования распределение напряженности электрического поля в фотонном кристалле:*
*а – при одном открытом канале, б – при обоих открытых каналах; d = 155 нм, a = 600 нм*

Для обобщенной количественной оценки оптимальных геометрических параметров фотонного кристалла построены графики зависимостей отношения напряженности электрического поля на выходах (*O*, output) и входах (*I*, input) оптического компаратора (рис. 7).




а б

*Рис. 7. Зависимость отношения интенсивностей сигналов на выходе и входе компаратора (O/I) от диаметра столбцов GaAs (а) и от периода компаратора (б)*

Важно отметить, что из-за переотражения излучения в фотонном кристалле при вводе излучения в один из каналов во втором канале тоже может наблюдаться нежелательное возрастание напряженности электрического поля. В связи с этим отношение интенсивностей сигналов в этом случае рассчитывалось как отношение разницы напряженностей на входах $O_1$ и $O_2$ и напряженности на входе $I_1$ (кривая с квадратными символами на рис. 7). Это значение должно быть максимально возможным, чтобы излучение проходило через волновод с минимальным затуханием и при этом практически не достигало второго выхода. Отношение выходных сигналов к входным при работе двух каналов должно быть минимальным, чтобы большая часть излучения, наоборот, затухала на пересечении двух волноводов (кривая с круглыми символами на рис. 7). Кривая с треугольными символами иллюстрирует разность в значениях отношения выходного и входного сигналов при подаче излучения в один и оба канала. Именно это значение должно быть

максимальным для достижения наилучших характеристик компаратора. Как видно на рис. 7,а, оптимальный диаметр столбцов фотонного кристалла для реализации на его основе оптического компаратора под излучение с длиной волны 1.3 мкм составляет 155 нм, в то время как период (расстояние между столбцами) – 600 нм (рис. 7,б).

**Заключение.** В данной работе проведено моделирование напряженности электрического поля в оптическом логическом компараторе на основе фотонного кристалла GaAs. Геометрические параметры фотонного кристалла сконфигурированы под излучение с длиной волны 1.3 мкм, находящейся во втором окне прозрачности оптического волокна. Проведено исследование влияния диаметра структурных элементов фотонного кристалла (столбцов GaAs) и расстояния между ними на степень затухания электромагнитного излучения, проходящего через волноводы, сформированные от входов к выходам компаратора. Установлено, что оптимальным значением диаметра столбцов GaAs для передачи излучения с длиной волны 1.3 мкм является 155 нм, а значение периода – 600 нм.

***Финансирование.*** *Работа выполнена при поддержке проекта Минобрнауки № FENW-2022-0034.*

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Yablonovitch E.* Inhibited Spontaneous Emission in Solid-State Physics and Electronics // Phys. Rev. Lett. – 1987. – Vol. 58, No. 20. – P. 2059-2062.
2. *Dyachenko P.N., Miklyaev Y.V., Dmitrienko V.E.* Three-dimensional photonic quasicrystal with a complete band gap // JETP Lett. – 2007. – Vol. 86, No. 4. – P. 240-243.
3. *Chutinan A., Kherani N.P., Zukotynski S.* High-efficiency photonic crystal solar cell architecture // Opt. Express. – 2009. – Vol. 17, No. 11. – P. 8871.
4. *Shekari Firouzjaei A., Salman Afghahi S., Ebrahimi Valmoozi A.-A.* Emerging Trends, Applications, and Fabrication Techniques in Photonic Crystal Technology // Recent Advances and Trends in Photonic Crystal Technology. – IntechOpen, 2024.
5. *Ветлужский А.Ю.* Волноводные устройства на основе линейных дефектов в металлических электромагнитных кристаллах // Журнал технической физики. – 2017. – Vol. 87, No. 1. – P. 150.
6. *Xavier S.C. et al.* Compact photonic crystal integrated circuit for all-optical logic operation // IET Optoelectron. – 2016. – Vol. 10, No. 4. – P. 142-147.
7. *Hassan S., Chack D., Pavesi L.* High extinction ratio thermo-optic based reconfigurable optical logic gates for programmable PICs // AIP Adv. – 2022. – Vol. 12, No. 5.
8. *Salmanpour A., Mohammadnejad S., Omran P.T.* All-optical photonic crystal NOT and OR logic gates using nonlinear Kerr effect and ring resonators // Opt. Quantum Electron. – 2015. – Vol. 47, No. 12. – P. 3689-3703.
9. *Bjarklev A., Chinlon Lin.* Applications of photonic crystal fibers in optical communications - What is in the future? // 2005 IEEE LEOS Annual Meeting Conference Proceedings. – IEEE, 2005. – P. 812-813.
10. *Welch D.F. et al.* High power, AlGaAs buried heterostructure lasers with flared waveguides // Appl. Phys. Lett. – 1987. – Vol. 50, No. 5. – P. 233-235.
11. *Babichev A.V. et al.* Heterostructures of Quantum-Cascade Lasers Based on Composite Active Regions // Bull. Russ. Acad. Sci. Phys. – 2023. – Vol. 87, No. 6. – P. 839-844.
12. *Tuchin V.V., Skibina J.S., Malinin A.V.* Photonic crystal fibers in biophotonics / ed. Popp J. – 2011. – P. 83110N.
13. *Tamer A.* Moniem. All-optical XNOR gate based on 2D photonic-crystal ring resonators // Quantum Electron. – 2017. – Vol. 47, No. 2. – P. 169-172.
14. *Sun Xiao-Wen et al.* Design and analysis of logic NOR, NAND and XNOR gates based on interference effect // Quantum Electron. – 2018. – Vol. 48, No. 2. – P. 178-183.
15. *Parandin F.* Ultra-compact terahertz all-optical logic comparator on GaAs photonic crystal platform // Opt. Laser Technol. – 2021. – Vol. 144. – P. 107399.
16. *Olyaee S., Naraghi A., Ahmadi V.* High sensitivity evanescent-field gas sensor based on modified photonic crystal fiber for gas condensate and air pollution monitoring // Optik (Stuttg). – 2014. – Vol. 125, No. 1. – P. 596-600.
17. *Brès C.-S. et al.* Supercontinuum in integrated photonics: generation, applications, challenges, and perspectives // Nanophotonics. – 2023. – Vol. 12, No. 7. – P. 1199-1244.

18. *Gorelik V.S. et al.* Three-dimensional quantum photonic crystals and quantum photonic glasses // Russ. J. Gen. Chem. – 2013. – Vol. 83, No. 11. – P. 2125-2131.
19. *Горбацевич А.А., Фриман А.В., Горелик В.С.* Двумерный гексагональный фотонный кристалл с новой геометрией элемента // Краткие сообщения по физике ФИАН. – 2014. – Т. 6. – С. 37-38.
20. *Драчев В.П.* Кремниевая фотоника: статус и перспективы развития // Матер. X Международного семинара по волоконным лазерам 2022. Институт автоматики и электрометрии СО РАН, 2022. – С. 200-201.

REFERENCES

1. *Yablonovitch E.* Inhibited Spontaneous Emission in Solid-State Physics and Electronics, *Phys. Rev. Lett.*, 1987, Vol. 58, No. 20, pp. 2059-2062.
2. *Dyachenko P.N., Miklyaev Y.V., Dmitrienko V.E.* Three-dimensional photonic quasicrystal with a complete band gap, *JETP Lett.*, 2007, Vol. 86, No. 4, pp. 240-243.
3. *Chutinan A., Kherani N.P., Zukotynski S.* High-efficiency photonic crystal solar cell architecture, *Opt. Express*., 2009, Vol. 17, No. 11, pp. 8871.
4. *Shekari Firouzjaei A., Salman Afghahi S., Ebrahimi Valmoozi A.-A.* Emerging Trends, Applications, and Fabrication Techniques in Photonic Crystal Technology, *Recent Advances and Trends in Photonic Crystal Technology*. IntechOpen, 2024.
5. *Vetluzhskiy A.Yu.* Volnovodnye ustroystva na osnove lineynykh defektov v metallicheskikh elektromagnitnykh kristallakh [Waveguide devices based on linear defects in metallic electromagnetic crystals], *Zhurnal tekhnicheskoy fiziki* [Journal of Technical Physics], 2017, Vol. 87, No. 1, pp. 150.
6. *Xavier S.C. et al.* Compact photonic crystal integrated circuit for all-optical logic operation, *IET Optoelectron,* 2016, Vol. 10, No. 4, pp. 142-147.
7. *Hassan S., Chack D., Pavesi L.* High extinction ratio thermo-optic based reconfigurable optical logic gates for programmable PICs, *AIP Adv*., 2022, Vol. 12, No. 5.
8. *Salmanpour A., Mohammadnejad S., Omran P.T.* All-optical photonic crystal NOT and OR logic gates using nonlinear Kerr effect and ring resonators, *Opt. Quantum Electron,* 2015, Vol. 47, No. 12, pp. 3689-3703.
9. *Bjarklev A., Chinlon Lin.* Applications of photonic crystal fibers in optical communications - What is in the future?, *2005 IEEE LEOS Annual Meeting Conference Proceedings*. IEEE, 2005, pp. 812-813.
10. *Welch D.F. et al.* High power, AlGaAs buried heterostructure lasers with flared waveguides, *Appl. Phys. Lett*., 1987, Vol. 50, No. 5, pp. 233-235.
11. *Babichev A.V. et al.* Heterostructures of Quantum-Cascade Lasers Based on Composite Active Regions, *Bull. Russ. Acad. Sci. Phys*., 2023, Vol. 87, No. 6, pp. 839-844.
12. *Tuchin V.V., Skibina J.S., Malinin A.V.* Photonic crystal fibers in biophotonics, ed. Popp J., 2011, pp. 83110N.
13. *Tamer A.* Moniem. All-optical XNOR gate based on 2D photonic-crystal ring resonators, *Quantum Electron,* 2017, Vol. 47, No. 2, pp. 169-172.
14. *Sun Xiao-Wen et al.* Design and analysis of logic NOR, NAND and XNOR gates based on interference effect, *Quantum Electron,* 2018, Vol. 48, No. 2, pp. 178-183.
15. *Parandin F.* Ultra-compact terahertz all-optical logic comparator on GaAs photonic crystal platform, *Opt. Laser Technol*., 2021, Vol. 144, pp. 107399.
16. *Olyaee S., Naraghi A., Ahmadi V.* High sensitivity evanescent-field gas sensor based on modified photonic crystal fiber for gas condensate and air pollution monitoring, *Optik (Stuttg),* 2014, Vol. 125, No. 1, pp. 596-600.
17. *Brès C.-S. et al.* Supercontinuum in integrated photonics: generation, applications, challenges, and perspectives, *Nanophotonics,* 2023, Vol. 12, No. 7, pp. 1199-1244.
18. *Gorelik V.S. et al.* Three-dimensional quantum photonic crystals and quantum photonic glasses, *Russ. J. Gen. Chem*., 2013, Vol. 83, No. 11, pp. 2125-2131.
19. *Gorbatsevich A.A., Friman A.V., Gorelik B.C.* Dvumernyy geksagonal'nyy fotonnyy kristall s novoy geometriey elementa [Two-dimensional hexagonal photonic crystal with a new element geometry], *Kratkie soobshcheniya po fizike FIAN* [Brief communications on physics of the Lebedev Physical Institute], 2014, Vol. 6, pp. 37-38.
20. *Drachev V.P.* Kremnievaya fotonika: status i perspektivy razvitiya [Silicon photonics: status and development prospects], *Mater. X Mezhdunarodnogo seminara po volokonnym lazeram 2022. Institut avtomatiki i elektrometrii SO RAN, 2022* [Proceedings of the X International Seminar on Fiber Lasers 2022. Institute of Automation and Electrometry SB RAS, 2022], pp. 200-201.

Статью рекомендовал к опубликованию д.т.н., профессор К.Е. Румянцев.

**Пленингер Максимилиан** – Южный федеральный университет; e-mail: pleninger@sfedu.ru; г. Таганрог, Россия; тел.: +79897471548; техник-проектировщик Лаборатории эпитаксиальных технологий.

**Балакирев Сергей Вячеславович** – e-mail: sbalakirev@sfedu.ru; тел.: +78634371611; к.т.н.; ведущий научный сотрудник Лаборатории эпитаксиальных технологий**.**

**Солодовник Максим Сергеевич** – e-mail: solodovnikms@sfedu.ru; тел.: +78634371611; к.т.н.; ведущий научный сотрудник Лаборатории эпитаксиальных технологий.

**Pleninger Maximilian** – Southern Federal University; e-mail: pleninger@sfedu.ru; Taganrog, Russia; phone: +79897471548; the design technician of Laboratory of Epitaxial Technologies.

**Balakirev Sergey Vyacheslavovich** – e-mail: sbalakirev@sfedu.ru; phone: +78634371611; cand. of eng. sc.; leading researcher of Laboratory of Epitaxial Technologies.

**Solodovnik Maxim Sergeevich** – e-mail: solodovnikms@sfedu.ru; phone: +78634371611; cand. of eng. sc.; leading researcher of Laboratory of Epitaxial Technologies.

УДК 004.032 DOI 10.18522/2311-3103-2024-5-194-204

**А.Н. Самойлов, Н.Е. Сергеев, С.М. Гушанский, В.С. Потапов**

# РАЗРАБОТКА И ИССЛЕДОВАНИЕ КВАНТОВОЙ ГРАФОВОЙ МОДЕЛИ ДЛЯ СЖАТИЯ И РЕКОНСТРУКЦИИ ИЗОБРАЖЕНИЙ

*Подробно рассматриваются методы и подходы к применению квантовых алгоритмов для решения задач оптимизации и обработки изображений. Особое внимание уделено квантовой приближённой оптимизации (КПО) и применению квантовых сетей для задач сжатия и реконструкции данных. КПО представляет собой гибридный алгоритм, который объединяет квантовые и классические вычислительные процессы, позволяя эффективно решать сложные комбинаторные задачи. Основой КПО являются параметризованные унитарные операции, которые подвергаются оптимизации в ходе итераций. Этот подход даёт возможность учитывать уникальные особенности квантовой природы информации, что в ряде случаев позволяет достичь более высокой производительности, чем при использовании исключительно классических методов. В процессе реализации КПО одним из главных препятствий остаётся проблема шума, который может возникать, например, при использовании CNOT-гейтов. В статье обсуждаются различные стратегии снижения уровня шума, что является важной задачей для обеспечения стабильности и повышения точности работы квантовых алгоритмов. Например, рассматриваются методы изоляции отдельных операций и коррекции ошибок, что позволяет минимизировать влияние шума на результаты вычислений и улучшить точность квантовой оптимизации. Авторы также предлагают графовую интерпретацию квантовых моделей, которая основана на применении тензорных сетей. Такой подход позволяет эффективно упрощать вычислительные графы, за счёт чего удаётся оптимизировать ресурсы, требуемые для выполнения сложных квантовых операций. Этот метод также демонстрирует высокую эффективность в задачах сжатия и восстановления изображений, что открывает новые перспективы для применения квантовых сетей в области обработки данных. В статье описывается структура квантовых сетей, включающая многослойные квантовые гейты, которые позволяют более глубоко и детализированно обрабатывать изображения, обеспечивая как эффективное сжатие, так и качественное восстановление данных. Также был проведён анализ различных типов квантовых гейтов, таких как Адамар, Паули-X, Паули-Y и T-гейты. Эти гейты играют ключевую роль в эффективности квантовых алгоритмов, так как каждый из них вносит свой вклад в квантовую динамику и в способ манипуляции квантовыми состояниями.*

*Моделирование; квантовый алгоритм; кубит; модель квантового компьютера; запутанность; суперпозиция; квантовый оператор.*

**A.N. Samoilov, N.E. Sergeev, S.M. Gushanskiy, V.S. Potapov**

## DEVELOPMENT AND RESEARCH OF A QUANTUM GRAPH MODEL FOR IMAGE COMPRESSION AND RECONSTRUCTION

*The article discusses in detail the methods and approaches to the application of quantum algorithms for solving optimization and image processing problems. Particular attention is paid to quantum approximate optimization (QAO) and the use of quantum networks for data compression and reconstruction problems. QAO is a hybrid algorithm that combines quantum and classical computational processes, allowing one to efficiently solve complex combinatorial problems. QAO is based on parameterized unitary operations that are optimized during iterations. This approach makes it possible to consider the unique features of the quantum nature of information, which in some cases allows achieving higher performance than when using exclusively classical methods. In the process of implementing QAO, one of the main obstacles remains the problem of noise, which can arise, for example, when using CNOT gates. The article discusses various strategies for reducing the noise level, which is an important task for ensuring the stability and improving the accuracy of quantum algorithms. For example, methods for isolating individual operations and correcting errors are considered, which allows one to minimize the impact of noise on the calculation results and improve the accuracy of quantum optimization. The authors also propose a graph interpretation of quantum models based on the use of tensor networks. This approach allows for efficient simplification of computational graphs, thereby optimizing the resources required to perform complex quantum operations. This method also demonstrates high efficiency in image compression and restoration tasks, which opens up new prospects for the application of quantum networks in data processing. The article describes the structure of quantum networks, including multilayer quantum gates, which allow for deeper and more detailed image processing, providing both efficient compression and high-quality data restoration. An analysis of various types of quantum gates, such as Hadamard, Pauli-X, Pauli-Y, and T-gates, was also conducted. These gates play a key role in the efficiency of quantum algorithms, since each of them contributes to quantum dynamics and the way quantum states are manipulated.*

*Modeling; quantum algorithm; qubit; model of a quantum computer; entanglement; superposition; quantum operator.*

**Введение.** Алгоритмы квантовой приближенной оптимизации [1] (КПО) – это гибридные методы, сочетающие квантовые и классические вычисления для решения комбинаторных оптимизационных задач. Алгоритм работает на основе специальных математических структур, таких как гамильтониан [2] задачи и гамильтониан смесителя. В процессе работы КПО применяются две параметризованные унитарные операции $U(\theta, \gamma)$ и $U(\theta, \beta)$, последовательно действующие на начальные квантовые состояния $U(\gamma) = \exp(-i\gamma)$ и $U(\beta) = \exp(-i\beta)$. Параметры этих процессов инициализируются случайным образом и обновляются после нескольких итераций с помощью классических методов оптимизации. Процесс повторяется с обновленными параметрами для приближения к оптимальному решению оптимизационной задачи.

В литературе также обсуждались методы уменьшения влияния шума на производительность КПО и других гибридных квантово-классических алгоритмов. Эти методы предполагают изменение гамильтониана или функции стоимости смесителя для уменьшения шума и увеличения скорости сходимости алгоритма. Важным аспектом улучшения производительности квантовых схем является уменьшение ошибок, вызванных CNOT-гейтами [3]. Эти затворы являются источником ошибок в современных квантовых устройствах, и методы оптимизации, снижающие их влияние на шум, играют важную роль в улучшении квантовых алгоритмов. Таким образом, КПО исследует различные аспекты оптимизации в квантовых вычислениях, и его развитие может привести к созданию более эффективных квантовых алгоритмов и, следовательно, является перспективным направлением для будущих исследований в области квантовых вычислительных технологий.

**1. Кодирование изображений в квантовые состояния.** Для изображений и общих классических данных матрица данных может быть преобразована в N-мерный массив векторов X [4]. Классические данные изображения x кодируются в амплитуды вероятно-

сти A квантовых состояний ψ. Подготовленное квантовое состояние ψ вводится в сеть сжатия U для получения квантового состояния в нижнем d-мерном пространстве. При этом измеряется выходное состояние U, вычисляется градиент параметров в соответствии с функцией потерь, и оптимизатор возвращает оптимальные параметры U. Квантовое состояние в d-мерном пространстве вводится в сеть реконструкции U и получается выходное состояние Ψ в многомерном пространстве; процесс обучения параметров для U такой же, как и для U. Вероятность B выходного состояния [5] в сети реконструкции преобразуется в классические данные изображения x для выполнения реконструкции изображения [6].

**2. Слияние изображений с использованием квантовых сетей.** Сжатие и восстановление изображений с помощью квантовых сетей [7] можно разделить на две независимые сети, которые можно представить как сеть квантового сжатия UC и сеть квантовой реконструкции UR. Целью сети сжатия является сжатие запутанных состояний в d-мерном гильбертовом пространстве с достаточно высокой точностью ориентации для достижения квантового сжатия состояний. Цель сети реконструкции – восстановить квантовые состояния в d-мерном гильбертовом пространстве в n-мерном гильбертовом пространстве, не приближая точность состояний к 100%. Процесс реконструкции может быть обратным процессу сжатия. В частности, сеть реконструкции UR может представлять собой комбинацию квантовых ворот сети сжатия, соединенных в обратном порядке, так что параметры сети необходимо заново изучать. Это связано с тем, что обратная сеть сжатия UC, $U^{-1}$, может быть напрямую использована в качестве сети реконструкции UR (UR = $U^{-1}$) только в том случае, если ошибки в сети сжатия малы. Если же ошибка не равна почти нулю, то целесообразнее использовать переобученную реконструированную сеть. Структуры схем являются гибкими, что облегчает проектирование сложных квантовых схем и обеспечивает масштабируемость квантовых вычислений. Идеальные многопортовые оптические интерферометры без потерь используются для преобразования между N-мерными векторными пространствами, которые могут быть описаны N × N унитариями.

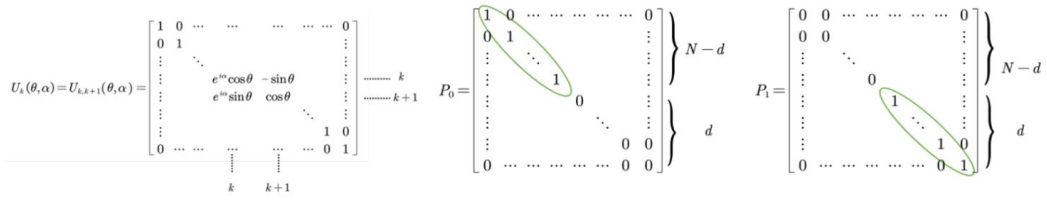


*Рис. 1. Квантовые гейты: uk, P1, P0. uk – квантовые гейты, P1 и P0 – проективные преобразования для сжатия*

Единые квантовые гейты [8] U представляют собой полное однородное преобразование всех кубитов, а $U_{(k,k+1)}$ – квантовое преобразование [9] k-го и векторного пространства. Для преобразования всех кубитов требуется интерферометр. Для обучения и обновления параметров θ используется алгоритм градиентного спуска [10]. Затем предлагается стратегия градиентного спуска для квантового сжатия и восстановления параметров сети. Непрерывное преобразование, состоящее из N 1-квантовых гейт, рассматривается как однослойное соединение квантовых гейтов U. Однако в реальной сети требуются многослойные квантовые ворота. В соответствии с предельным определением производной, производная параметров сети [11] может быть определена следующим образом.

$$U = U_{(1,2)}U_{(2,3)}U_{(3,4)}[...]U_{(k,k+1)}$$

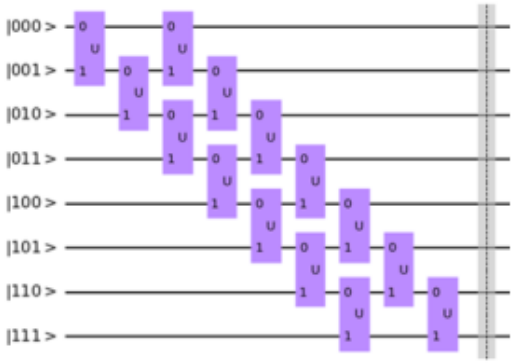



*Рис. 2. Структура квантовой сети. В каждом слое этой квантовой сети* $U_{(k,k+1)}$ *квантовых гейтов соединены в порядке серого кода; количество квантовых гейтов U в слое равно N - 1*

Основная задача квантовой сети – узнать параметр отражения θ всех допустимых квантовых ворот. Конкретные наборы данных выбираются для онлайн-обучения с помощью моделирования, а реальная физическая реализация может быть установлена в подходящем формате.

**3. Графовая интерпретация квантовой модели для исключения гейтов в квантовых схемах.** Разработанный алгоритм основан на тензорном сокращении сети, в которой ширина графа [12] будет доминирующим фактором в определении временной и вычислительной сложности. Алгоритм подразумевает выполнение следующего набора этапов:

1. Построить упрощенный неориентированный граф;

2. Разделить всю задачу на многие подзадачи, сохранив время работы каждой подзадачи как можно короче;

Основная идея состоит в том, чтобы удалить самые дорогостоящие узлы, распараллелив их значения.

3. Используем эту оценку, чтобы определить эффективный порядок устранения остальных узлов в каждой подзадаче.

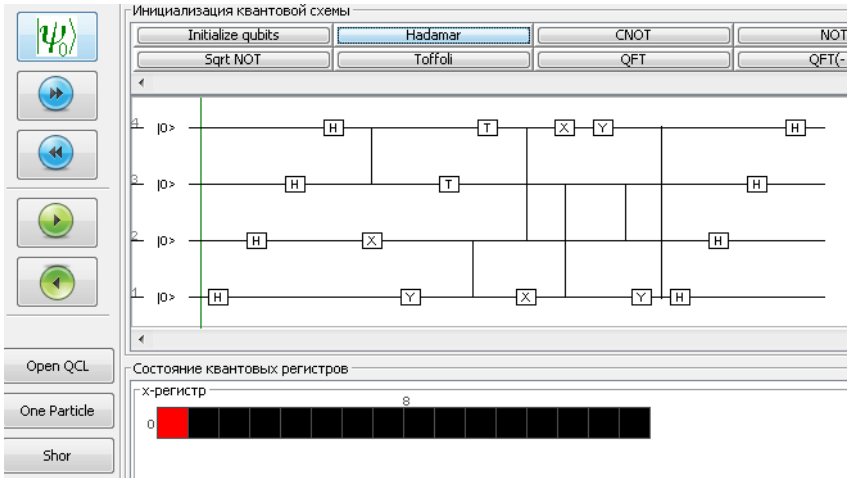



*Рис. 3. Пример схемы C, которую оцениваем с использованием графовой модели*

КВС (рис. 3) основан на сложной неориентированной графовой модели. Это, по существу, вариант тензорной сети [13], которая также использует диагональные гейты.

Рассмотрим неориентированную графовую модель, которая также используется в алгоритме. Для любой квантовой схемы C амплитуда конкретной битовой строки x, $\langle x|C|0...0\rangle$, задается формулой

$$\langle x|C|0...0\rangle = \langle x|C_\partial ...C_2C_1|0...0\rangle, \quad (1)$$

где $C_d...C_2C_1$ – унитарные матрицы, соответствующие тактовым циклам 1, 2 и d, соответственно. Можно дополнительно разложить формулу как

$$\langle x|C|0...0\rangle = \langle x|C_d...C_2C_1|0...0\rangle = \sum_{i_1,i_2,...,i_{d-1}\in\{0,1\}^N} \langle x|C_d|i_{d-1}\rangle ...\langle i_2|C_2|i_1\rangle\langle i_1|C_1|0...0\rangle. \quad (2)$$

Например, рассмотрим схему с N = 4 кубитов и глубиной d = 8, как показано на рис. 3. Определенное количество и расположение дублирующих квантовых гейтов в некоторых ветвях [14] необходимо для демонстрации процедуры исключения переменных, вершин графа (элементов квантовой схемы) при выполнении редукции. Стоит также отметить, что степень редукции графа (элементов квантовой схемы) напрямую зависит от наполнения квантовой схемы и расположения ее гейтов.

Квантовые алгоритмы и квантово-физические процессы являются довольно перспективной и быстрорастущей областью научных исследований. Во многих из алгоритмов используется практически один и тот же набор квантовых гейтов для реализации их работы лишь с некоторыми вариациями гейтов. В работе отражен конкретный набор квантовых гейтов, который проиллюстрирован в квантовой схеме [15] (гейты Адамара, X, Y и T). Этот набор является необходимым и достаточным благодаря произведенному подробному анализу базы существующих квантовых алгоритмов и процессов. Это помогло выявить процентное соотношение использования тех или иных квантовых гейтов, участвующих в работе конкретных квантовых алгоритмов [16] или процессов. Not (10%), CNot (15%), CCNot (10%), Hadamard (75%), Паули-X гейт (35%), Паули-Y гейт (30%), Паули-Z гейт (20%), √Not (5%), Swap (5%), √SWAP (5%), CSWAP (10%), XX (5%), гейт Дойча (10%), фазовый сдвиг (5%), T (20%), квантовое преобразование Фурье (10%), инверсивное квантовое преобразование Фурье (5%), CZ (10%), C (квантовый оракул) (10%). Можно сделать вывод, что гейты Адамара, X, Y и T имеют наибольший показатель использования (20% и выше) среди квантовых схем различных квантовых алгоритмов и процессов.

Под наполнением квантовой схемы подразумевается процентное соотношение заполненности всех линий, из которых состоят квантовые схемы. Вычислим наполненность квантовой схемы. Данная схема имеет 4 линии и 64 мест для заполнения их квантовыми гейтами, так как каждая линия имеет 16 позиций для квантовых гейтов. Следовательно, процент наполнения схемы составляет (100% * 16) / 64 = 25%. Рассмотрим $C = C_8...C_2C_1$ как

$$C_1 = H_1 \otimes H_2 \otimes H_3 \otimes H_4,\ C_2 = CZ_{1,2} \otimes \sqrt{X_3} \otimes \sqrt{Y_4},\ C_3 = T_1 \otimes T_2 \otimes CZ_{3,4},$$
$$C_4 = CZ_{1,3} \otimes I_2 \otimes \sqrt{X_4}, C_5 = CZ_{2,4} \otimes I_3 \otimes \sqrt{X_1},\ C_6 = CZ_{2,3} \otimes \sqrt{Y_1} \otimes \sqrt{Y_4},$$
$$C_7 = CZ_{1,4} \otimes I_{2,3},\ C_8 = H_1 \otimes H_2 \otimes H_3 \otimes H_4.$$

Затем можно раскрыть (70) как

$$\sum_{i_1,i_2,...,i_7\in\{0,1\}^4} \langle x|H_1 \otimes H_2 \otimes H_3 \otimes H_4|i_7\rangle\langle i_7|CZ_{1,4} \otimes I_{2,3}|i_6\rangle$$
$$\cdot\langle i_6|CZ_{2,3} \otimes \sqrt{Y_1} \otimes \sqrt{Y_4}|i_5\rangle\langle i_5|CZ_{2,4} \otimes I_3 \otimes \sqrt{X_1}|i_4\rangle\langle i_4|CZ_{1,3} \otimes I_2 \otimes \sqrt{X_4}|i_3\rangle \quad (3)$$
$$\langle i_3|T_1 \otimes T_2 \otimes CZ_{3,4}|i_2\rangle\langle i_2|CZ_{1,2} \otimes \sqrt{X_3} \otimes \sqrt{Y_4}|i_1\rangle\langle i_1|H_1 \otimes H_2 \otimes H_3 \otimes H_4|0000\rangle$$

Рассмотрим суммирование в (3), выполняющееся по семи 4-битным строкам $i_1,...,i_7$. Поскольку T и CZ являются диагональными матрицами [17], члены суммирования которых появятся только в том случае, если $i_1^{(1,2)} = i_2^{(1,2)}$, $i_2 = i_3$, $i_3^{(1,2,3)} = i_4^{(1,2,3)}$, $i_4^{(2,3,4)} = i_5^{(2,3,4)}$, $i_5^{(2,3)} = i_6^{(2,3)}$ и $i_6 = i_7$. Рассматривая диагональные гейты таким образом, резко сокращается общее число членов от $2^{28}$ до $2^{10}$.

Сформулируем вышеприведенную процедуру на языке неориентированных графовых моделей. Учитывая последовательности индексов в (2), построим граф G, где две вершины связаны ребром, если на них действует оператор.

Не все вершины графа подлежат редукции. Граф можно упростить, если тензорные операторы оказываются диагональными. Например, если узлы u и v связаны однокубитовым [18] диагональным гейтом, то один член в интеграле траектории Фейнмана может остаться в графе только тогда, когда соответствующая маркировка, которую мы выбираем, удовлетворяет u = v. Поэтому два узла u и v можно объединить вместе. Аналогично, тензоры на левой стороне (выделены фоном) соответствуют своим графовым интерпретациям [19] справа.

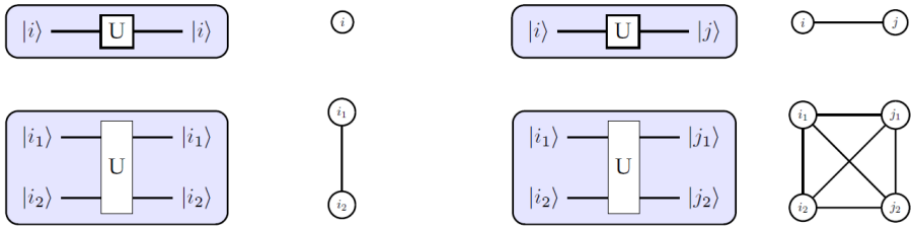



*Рис. 4. Одно- и двухкубитные диагональные и недиагональные гейты и их графовые интерпретации*

Алгоритм оценки суммы в уравнении 2 продолжается путем устранения по одной вершине за один такт работы алгоритма. Обычно этот процесс приводит к получению тензоров ранга [20], большего, чем 2. Для устранения двоичной вершине $v = i_k^j$ из (2):

1. Найти множество тензоров $T_v = \{\tau \mid v \to \tau\}$ и множество вершин $V_v = \bigcup_{\tau \in T_v} \{v' \mid v' \to \tau\}$

2. Умножить все тензоры $\tau \to T_v$ для получения нового тензора $\sigma$ ранга $|V_v|$ и проиндексировать вершины в $|V_v|$;

3. Просуммировать $\sigma$ над индексом, соответствующим v, чтобы получить $\sigma'$;

4. Удалите вершину v и все тензоры в $T_v$ из суммирования, а затем добавить новый тензор $\sigma'$. Обновить неориентированный граф, соединяющий любые две соседние вершины v, а затем удалить v.

Чтобы оценить всю сумму, повторим описанные выше шаги, чтобы устранить все вершины в любом удобном порядке. Когда все вершины удаляются, мы получаем тензор ранга 0 (т.е. комплексное число), который является точным значением амплитуды.

Существует еще более простой метод для оценки (2), который заключается лишь в разделении суммы на части. Можем просто выбрать любую вершину v и оценить суммирование дважды, один раз со значением v, зафиксированным на 0, и один раз со значением v, зафиксированным на 1, и затем объединить результаты. Подобно устранению вершины, фиксация значения вершины также удаляет ее из суммирования. В модели неориентированного графа фиксация значения вершины приводит к удалению соответствующей вершины вместе со всеми его краями.

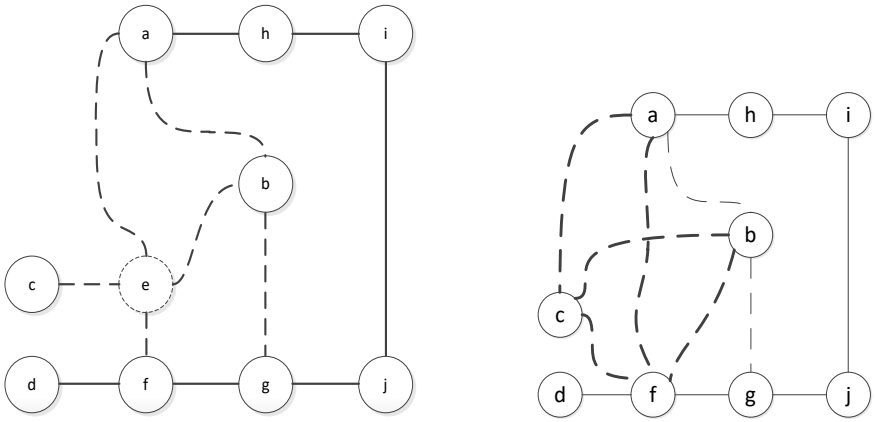



*Рис. 5. Распараллеливание путем фиксации значения узла*

Начальная неориентированная графовая модель показана слева. Фиксация значения узла «c» не помогает значительно упростить граф. Вместо этого значение узла «e» значительно упрощает граф. С другой стороны, оценка неориентированного графа с использованием только этой стратегии была бы крайне неэффективной по сравнению с реализованным алгоритмом. Общая стратегия – удалить t вершин графа таким образом, чтобы, по существу, разделить задачу оценки графа на $2^t$ подзадач, представляющих собой набор вершин неориентированной графовой модели.

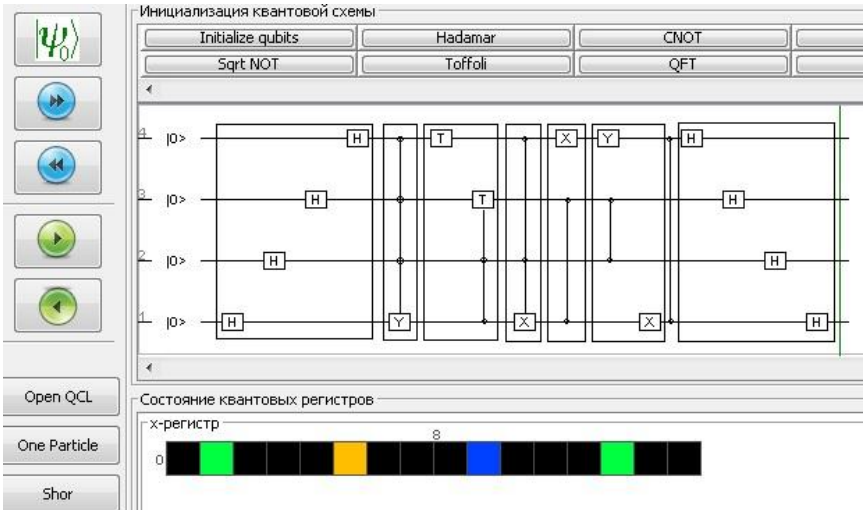
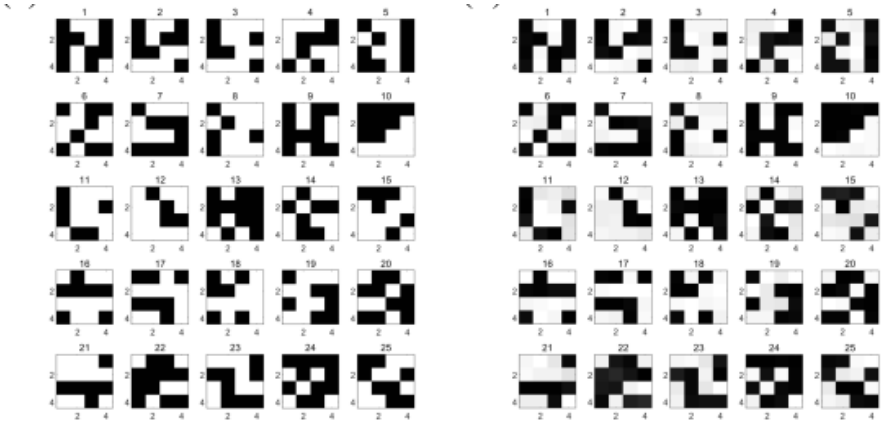



*Рис. 6. Квантовая схема C, полученная путем редукции графовой модели*

Затем происходит выполнение каждой подзадачи, используя базовый алгоритм. Конечно, это действует только в том случае, если стоимость оценки полученного графа в значительной степени ниже стоимости оценки исходного графа. Далее можно произвести оптимизацию вычислительных процессов в квантовой области с помощью трансляции редуцированного графа в квантовую схему. Каждый такт работы квантовой схемы выделен в отдельную область.

**Набор данных и настройка параметров**. Используем Matlab для моделирования процесса квантового алгоритма [21], выполняемого на классическом компьютере. Сначала выбираем 25 бинарных изображений и задаем основные параметры сети. В процессе обучения настраиваем параметры сети в соответствии с эффектами сжатия и реконструкции изображений, включая количество слоев квантовой сети и скорость обучения [22].

Наконец, минимальные значения LC и LR почти равны 0, что означает, что обучение носит практический характер. При этом точность реконструкции обучения достигает 97,75%. Данные о сжатии и реконструкции могут быстро сходиться к целевому значению. На рис. показан процесс обновления θ, где конечные параметры стабильны, градиент падает до 0, а параметр θ находится в пределах [0, 2π]. Используя ту же структуру масштабной сети (словарь 16*16) для того же набора данных, сравнили CSC на основе алгоритмов SVD с алгоритмом на основе QN.



а б

*Рис. 7. Процесс обучения сжатию и реконструкции изображений: а – входное двоичное изображение в размерности 4x4 (x, 0, 1, M = 25). Эти изображения необходимо закодировать как амплитуды 3-кубитных состояний. б – результаты реконструкции изображения*

В частности, в CSC можем использовать вектор разреженного кодирования [23] s и словарь D для выражения входного сигнала y, обозначаемого как y = Ds. Можем увидеть сравнение потерь обучения, как показано на рис. 5. Таким образом, наш алгоритм на основе QN в определенной степени имеет лучшее квантовое превосходство, чем алгоритм на основе CSC. Аналогично, по сравнению с другими квантовыми алгоритмами, преимущества предложенной нами структуры квантовой сети могут быть существенно отражены. На рис. 8 изображена разработанная модель квантового вычислительного устройства и результат выполнения квантового алгоритма преобразования набора пикселей. Данная программная разработка является offline десктопной вычислительной системой с открытой архитектурой, однако в дальнейшем планируется ее трансляция в online режим. Процесс измерения заключается в разыгрывании числа случайного характера из промежутка [0, 1]. При попадании значения в интервал $[0, |c_1|^2]$ выходным значением измерительного процесса является базисное состояние бра вектора 0, в противном случае (интервал $[|c_1|^2, 1]$) – квантовое состояние бра вектора 1.

В новейших квантовых компьютерных технологиях возможно хорошо контролировать и управлять только ограниченным числом кубитов [24]. Тогда как, с другой стороны, реальная проблема оптического распознавания символов в квантовом методе опорных векторов требует десятки кубитов, что не может быть осуществлено в настоящее время. Таким образом, ограничим проблему случаем с минимальными затратами, в котором только два варианта («6» либо «9») находятся в списке и только два свойства (горизонтальное либо вертикальное положение, определяемое в дальнейшем) составляют задачу. Это позволяет продемонстрировать данный алгоритм квантового искусственного интеллекта на основе 4-кубитного момента ядра квантового процессора при комнатной температуре. В разработанном алгоритме используются квантовые принципы, такие как суперпозиция квантовых состояний вычислительной системы, запутанность квантовых состояний и преобразование классического изображения в квантовое состояние путем кодирования цветовой палитры пиксельного набора в

рамках комплексных амплитудных квантовых состояний. Разработанная модель квантового вычислителя обладает рядом важных характеристик: модульность, библиотечный набор, исходное количество гейтов, возможность добавления гейтов, редуцированная матрица состояний, поддержка платформ Windows/Linux, редактор квантовой схемы, задание входных значение кубит, числовой вывод вероятностей/амплитуд кубита [25], таблица цветов вероятностей/амплитуд состояний кубит, автоматический режим, пошаговый режим, моделирование физических процессов, матричное математическое ядро, нет ограничения на количество кубит, сохранение/загрузка программы/схемы/алгоритма, сохранение результатов вычислений, загрузка и продолжение вычислений, открытая архитектура, отработанные примеры, документация. Разработка архитектуры вычислителя подразумевает под собой разработку структуры программы, которая включает разбиение структуры на программные компоненты и разработку схемы взаимодействия между этими компонентами с помощью доступных снаружи свойств и методов этих компонентов.

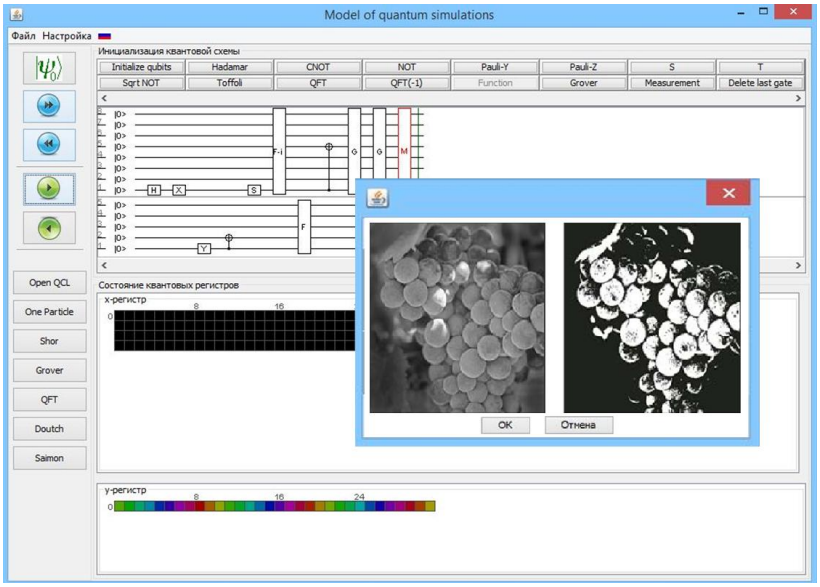



*Рис. 8. Модель квантового вычислительного устройства*

**Заключение.** В данной работе на основе классического алгоритма сжатия и реконструкции изображений создан соответствующий квантовый алгоритм, использующий квантовые состояния в качестве носителей информации об изображении. Сжатие и реконструкция изображения осуществляется с помощью квантовой сети сжатия и реконструкции. Реконструированные квантовые состояния [8] разлагаются на классическую информацию, и, наконец, выполняется сжатие и реконструкция классической информации изображения. Результаты моделирования полностью подтверждают эффективность и превосходство алгоритма сжатия и реконструкции изображений с помощью квантовых сетей: Реальная квантовая сеть с $\alpha = 0$ все еще имеет определенные ограничения для квантовых задач и может решать только реальные задачи сжатия и реконструкции информации. Поэтому в будущем необходимо поддерживать фазовый параметр $\alpha$ квантовых гейтов и строить полноценные сложные квантовые сети, пригодные для решения более разнообразных квантовых задач. Аналогично, ожидается, что построение сложных квантовых сетей обеспечит прямое решение проблемы сжатия и реконструкции известных или неизвестных квантовых состояний. В будущем классические приложения квантовых алгоритмов будут становиться все более распространенными, а алгоритмы сжатия и реконструкции изображений на основе квантовых сетей получат потенциальное практическое применение в квантовом зрении. Алгоритмы на основе qN могут быть вычисли-

тельно применены к общей квантовой визуализации, основанной на проектировании оптических квантовых схем. Алгоритм на основе qN – это алгоритм квантовой визуализации, основанный на разработке оптических квантовых схем. Представлены результаты реализации и моделирования с использованием квантового компьютера.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК


1. *Farhi E., Goldstone J., and Gutmann S.* A quantum approximate optimization algorithm, *arXiv preprint arXiv:1411.4028,* 2014.
2. *Cerezo M., Arrasmith A., Babbush R., Benjamin S., Endo S., Fujii K., McClean J., Mitarai K., Yuan X., Cincio L., et al.* Variational quantum algorithms, *arXiv preprint arXiv:2012.09265*, 2020.
3. *Linke N.M., Gutierrez M., Landsman K.A., et al.* Fault-tolerant quantum error detection, *Science Advances*, 2017, 3(10): e1701074. Available from: https://doi.org/10.1126/ sciadv.1701074.
4. *Vuillot C.* Is error detection helpful on IBM 5q chips?, *Quantum Information and Computation,* 2018, Vol. 18, No. 11-12, pp. 0949-0964.
5. *Barron G. and Wood C.* Measurement error mitigation for variational quantum algorithms, *arXiv preprint arXiv:2010.08520*, 2020.
6. *Endo S., Benjamin S., and Li Y*. Practical quantum error mitigation for near-future applications, *Physical Review X*, 2018, 8 (3):031027.
7. *Endo S., Cai Z., Benjamin S., and Yuan X.* Hybrid quantum-classical algorithms and quantum error mitigation, *Journal of the Physical Society of Japan*, 2021, 90 (3):032001.
8. *Zhu L., Tang H., Barron G., Calderon-Vargas F., Mayhall N., Barnes E., and Economou S.* An adaptive quantum approximate optimization algorithm for solving combinatorial problems on a quantum computer, arXiv preprint arXiv:2005.10258, 2020.
9. *Larkin J., Jonsson M., Justice D., and Guerreschi G.* Evaluation of quantum approximate optimization algorithm based on the approximation ratio of single samples, *arXiv e-prints, pages arXiv–2006*, 2020;
10. *Barkoutsos P., Nannicini G., Robert A., Tavernelli I., and Woerner S.* Improving variational quantum optimization using cvar, *Quantum*, 2020, 4:256.
11. *Harper R, Flammia ST.* Fault-tolerant logical gates in the IBM quantum experience, *Phys Rev Lett.,* 2019, 122:080504. Available from: https://link.aps.org/doi/10.1103/ PhysRevLett.122.080504,
12. *Hales L., Hallgren S.* An improved quantum Fourier transform algorithm and applications, *Proceedings of the 41st Annual Symposium on Foundations of Computer Science, November 12–14, 2000*, pp. 515.
13. *Guzik V., Gushanskiy S., Polenov M., Potapov V.* Complexity Estimation of Quantum Al-gorithms Using Entanglement Properties, *16th International Multidisciplinary Scientific GeoConference, Bulgaria, 2016,* pp. 20-26.
14. *Guzik V., Gushanskiy S., Polenov M., Potapov V.* Models of a quantum computer, their characteristics and analysis, *9th International Conference on Application of Information and Communication Technologies (AICT)*. Institute of Electrical and Electronics Engineers, 2015, pp. 583-587.
15. *Collier D.* The Comparative Method, In: Finifter A.W. (ed.), *Political Sciences: The State of the Discipline II*. American Science Association. Washington, DC, 1993, pp. 105-119.
16. *Olukotun K.* Chip Multiprocessor Architecture – Techniques to Improve Throughput and Latency. Morgan and Claypool Publishers, San Rafael, 2007.
17. *Raedt K.D., Michielsen K., De Raedt H., Trieu B., Arnold G., Marcus Richter, Th Lip-pert, Watanabe H., and Ito N.* Massively parallel quantum computer simulator, *Computer Physics Communications*, 176, pp. 121-136.
18. *Williams C.P.* Explorations in Quantum Computing, *Texts in Computer Science, Chapter 2. Quantum Gates*. Springer, 2011, pp. 51-122.
19. *Potapov V., Gushanskiy S., Guzik V., Polenov M.* The Computational Structure of the Quantum Computer Simulator and Its Performance Evaluation, *Software Engineering Perspectives and Application in Intelligent Systems. Advances in Intelligent Systems and Computing*. Springer, 2019, Vol. 763, pp. 198-207.
20. *Rahman M. and Kaykobad M.* Complexities of some interesting problems on spanning trees, *Information Processing Letters*, 2005, 94 (2), pp. 93-97.
21. *Bennett C.H., Shor P.W., Smolin J.A., Thapliyal A.V*. Entanglement-assisted Capacity of a Quantum Channel and the Reverse Shannon Theorem, *IEEE Transactions on Information Theory*, 2002, 48, 2637.
22. *Hector A. et al.* Qiskit: An open-source framework for quantum computing, 2019.
23. *Milner R.G.* A Short History of Spin, *In: Contribution to the XV International Workshop on Polarized Sources, Targets, and Polarimetry. Charlottesville, Virginia, USA, September 9-13, 2013*. arXiv:1311.5016, 2013.

24. *Boneh D., Zhandry M.* Quantum-secure message authentication codes, *Proceedings of Eurocrypt*, 2013, pp. 592-608.
25. *Potapov V., Gushansky S., Guzik V., Polenov M.* Architecture and Software Implementation of a Quantum Computer Model, *Advances in Intelligent Systems and Computing*. Springer, 2016, Vol. 465, pp. 59-68.

Статью рекомендовал к опубликованию д.ф.-м.н., профессор Г.В. Куповых.

**Самойлов Алексей Николаевич** – Южный федеральный университет; e-mail: asamoylov@sfedu.ru; г. Таганрог, Россия; тел.: 88634371656; кафедра вычислительной техники; к.т.н.; доцент.

**Гушанский Сергей Михайлович** – e-mail: smgushanskiy@sfedu.ru; тел.: 88634371656; кафедра вычислительной техники; к.т.н.; доцент.

**Сергеев Николай Евгеньевич** – e-mail: nesergeev@sfedu.ru; тел.: 88634371656; кафедра вычислительной техники; д.т.н.; профессор.

**Потапов Виктор Сергеевич** – e-mail: vitya-potapov@rambler.ru; тел.: 88634371656; кафедра вычислительной техники; ассистент.

**Samoilov Alexey Nikolaevich** – Southern Federal University; e-mail: asamoylov@sfedu.ru; Taganrog, Russia; phone: +78634371656; the Department of Computer Engineering; cand. of eng. sc.; associate professor.

**Gushanskiy Sergey Mikhailovich** – e-mail: smgushanskiy@sfedu.ru; phone: +78634371656; the Department of Computer Engineering; cand. of eng. sc.; associate professor.

**Sergeev Nikolay Evgenievich** – e-mail: nesergeev@sfedu.ru;phone: +78634371656; the Department of Computer Engineering; dr. of eng. sc., professor.

**Potapov Victor Sergeevich** – e-mail: vitya-potapov@rambler.ru; phone: +78634371656; the Department of Computer Engineering; assistant.

# Раздел III. Электроника, приборостроение и радиотехника

УДК 621.396.677 DOI 10.18522/2311-3103-2024-5-205-214

**И.Н. Бобков, Ю.В. Юханов**

## ДВУХПОЛЯРИЗАЦИОННАЯ АНТЕННАЯ РЕШЕТКА ВИВАЛЬДИ С УМЕНЬШЕННОЙ ВЫСОТОЙ ПРОФИЛЯ

*Исследован элемент плоской антенной решетки Вивальди, предназначенной для работы на двух линейных поляризациях. Излучатели антенной решетки представляют диэлектрические подложки с двусторонней металлизацией и состоят из расширяющегося щелевого раскрыва, распределенного симметрирующего трансформатора и короткого отрезка микрополосковой линии. При этом длина симметрирующего трансформатора уменьшена за счёт придания ему формы синусоиды и, таким образом, уменьшен продольный размер излучателей и высота профиля всей антенной решетки. Соединение соседних излучателей осуществляется при помощи металлических столбиков, расположенных на металлическом экране. Представлены результаты численного исследования характеристик согласования и излучения элементарной ячейки с периодическими граничными условиями на гранях. Показано, что несмотря на уменьшение длины излучателей Вивальди, за счёт миниатюризации распределенного симметрирующего трансформатора, в антенной решетке не происходит сужения полосы рабочих частот. Расчётный коэффициент усиления элементарной ячейки близок к теоретически достижимому коэффициенту направленного действия апертуры такой же площади, что и элементарная ячейка. Коэффициент полезного действия при излучении по нормали не опускается ниже 75% во всей полосе рабочих частот. Проведенное исследование характеристик излучения при сканировании луча диаграммы направленности в E-, H- и D-плоскостях показало возможность отклонения луча на 60° без появления эффекта «ослепления» антенной решетки. Установлено влияние развязки между ближайшими ортогональными элементами антенной решетки на КПД при сканировании луча в диагональной плоскости. Представленные результаты исследования кросс-поляризационных характеристик элемента при отклонении луча на угол 45° в D-плоскости показывают, что коэффициент усиления элементарной ячейки на кросс-поляризации меньше коэффициента усиления на ко-поляризации на значение от 6 до 15 дБ. Полоса рабочих частот, определяемая по уровню КСВН≤3, составила от 915 до 7500 МГц.*

*Антенные решетки; антенны Вивальди; направленные антенны; излучатели; щелевой раскрыв; распределенный симметрирующий трансформатор.*

**I.N. Bobkov, Y.V. Yukhanov**

## A DUAL-POLARIZED TAPERED SLOT ANTENNA ARRAY WITH REDUCED PROFILE HEIGHT

*An element of a planar Vivaldi antenna array designed to operate on two linear polarizations is considered. The antenna array element is made of a dielectric substrate with double-sided metallization and consist of a tapered slot, a balun and a short section of microstrip line. At the same time, the length of the balun is reduced by making it sine-shaped and, thus, the longitudinal size of the Vivaldi element and the profile height of the entire antenna array are reduced. The connection between adjacent elements of the antenna array is carried out using metal posts placed on a metal screen. The results of a numerical study of the matching and radiation characteristics of a unit-cell with periodic boundary conditions on the faces are presented. It is shown that despite the reduction of the length of the Vivaldi antennas, due to the miniaturization of the balun, there is no narrowing of the operating frequency band in the antenna array. The calculated gain of the unit-cell is close to the theoretically achievable directivity of the aperture of the same area as the unit-cell. The broadside radiation efficiency does not fall below 75% over the entire*

*operating frequency range. A study of the radiation characteristics when scanning the beam of the radiation pattern in the E-, H- and D-planes showed the possibility of deflecting the beam by an 60° angle without the appearance of antenna array blinding effects. The effect of small isolation between the nearest orthogonal elements of the antenna array on the matching at the input of the elements when scanning a beam in a diagonal plane is shown. The presented results of a study of the cross-polarization characteristics of the element when the beam is deflected at an angle of 45° in the D-plane show that the cross-polarization gain of the unit-cell is less than the co-polarization gain by an amount from 6 to 15 dB. The operating frequency band, determined by the VSWR≤3 level, ranged from 915 to 7500 MHz.*

*Antenna arrays; Vivaldi antennas; directional antennas; profile height; unit-cell.*

**Введение.** Антенные решетки (АР) Вивальди [1] обладают рядом достоинств, среди которых особо выделяется возможность работы в сверхширокой полосе частот. Однако для обеспечения сверхширокополосного согласования на входе элементов таких АР необходимо увеличивать продольные размеры излучателей [2].

При синтезе плоских АР, количество элементов в которых может исчисляться десятками тысяч [3], применение электрически длинных антенн увеличивает расход материалов, утяжеляет и усложняет конструкцию АР и вспомогательных несущих деталей. Другим менее явным следствием применения электрически длинных элементов является высокий уровень кросс-поляризации при сканировании луча вне основных плоскостей [4, 5].

Проблеме высокого уровня кросс-поляризации в сверхширокополосных АР Вивальди в последние годы уделяется значительное внимание. Предложены улучшенные конструкции излучателей [5–9] и найдены способы улучшения кросс-поляризационных характеристик плоских АР линейной поляризации [10, 11].

В то же время работы, посвященные таким прикладным вопросам, как уменьшение электрических размеров (прежде всего высоты профиля) и массы АР Вивальди, широко не представлены. В названиях опубликованных статей и докладов конференций встречается слово «компактный» по отношению к АР Вивальди, но либо описанные решения не являются сверхширокополосными [12, 13], либо являются компактными лишь в смысле применения малого количества элементов (которые при этом являются электрически длинными) [14].

Представленные в литературе действительно низкопрофильные и при этом сверхширокополосные плоские АР Вивальди либо подразумевают наличие побочных главных максимумов диаграммы направленности (ДН) (т.е. являются сверхширокополосными за счёт диапазона, расположенного выше частоты $f_h \approx c/2d$, где $c$ – скорость света в вакууме, $d$ – шаг АР) [15], либо работают за счёт обеспечения сильной связи и противофазного питания между соседними элементами, и скорее представляют собой АР сильно связанных диполей, чем традиционные АР Вивальди [16].

В настоящей работе исследованы характеристики элементарной ячейки двухполяризационной АР антиподальных [17] излучателей Вивальди, в которой продольные размеры излучателей уменьшены за счет миниатюризации распределенного симметрирующего трансформатора [18, 19]. Таким образом достигнуто уменьшение высоты профиля всей антенной решетки. В качестве прототипа выбрана плоская АР Вивальди линейной поляризации, исследованная в [9], где подробно описано построение синусоидального симметрирующего трансформатора и влияние его геометрических параметров на характеристики АР.

**Модель элементарной ячейки.** Общий вид элементарной ячейки двухполяризационной антенной решетки Вивальди с уменьшенной высотой профиля представлен на рис. 1. На верхней и нижней поверхностях элементарной ячейки заданы граничные условия (ГУ) излучения, периодические ГУ заданы на боковых поверхностях А, А’ и Б, Б’. Излучатели выполнены на диэлектрической подложке с $\varepsilon_r$=3.55, tan(δ)=0.0027 толщиной 508 мкм. Шаг АР составляет 20 мм, высота излучателей над поверхностью металлического экрана $h$=60 мм. Питание осуществляется при помощи коаксиальных соединителей типа IX.

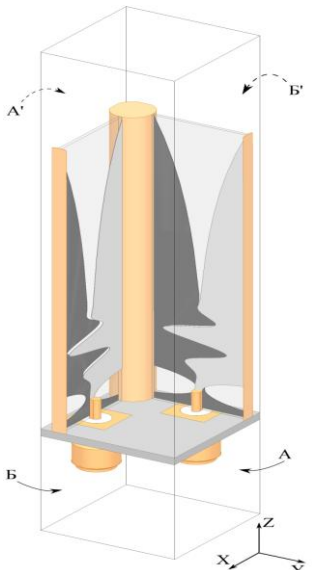



*Рис. 1. Общий вид элементарной ячейки двухполяризационной АР Вивальди с уменьшенной высотой профиля*

Конструкция отличается от традиционной антиподальной [17] формы симметрирующего трансформатора. Как показано в [9], применение симметрирующего трансформатора в форме синусоиды позволяет осуществлять более плавную трансформацию сопротивления на пути к раскрыву и тем самым сместить нижнюю границу рабочей полосы в область более низких частот. Этим расширением полосы частот можно пожертвовать для уменьшения высоты профиля АР [19], поскольку отношение высоты излучателя АР Вивальди к его ширине находится в прямой зависимости от ширины полосы рабочих частот АР [5].

Таким образом за счёт применения синусоидального симметрирующего трансформатора возможна разработка АР Вивальди с уменьшенной высотой профиля и такой же шириной полосы рабочих частот, что и у полноразмерных аналогов [19].

**Результаты численного исследования.** В программном обеспечении Ansys HFSS проведено численное исследование характеристик элементарной ячейки, изображенной на рис. 1.

На рис. 2 приведены рассчитанные зависимости КСВН на входе одного излучателя элементарной ячейки от частоты. Для сравнения приводятся ещё две зависимости КСВН на входе излучателей элементарных ячеек без синусоидальных симметрирующих трансформаторов: высотой 60 и 79 мм. В качестве верхней границы полосы рабочих частот принимается 7.5 ГГц. На этой частоте шаг АР начинает превышать половину длины волны и возможно появление побочных главных максимумов ДН [9].

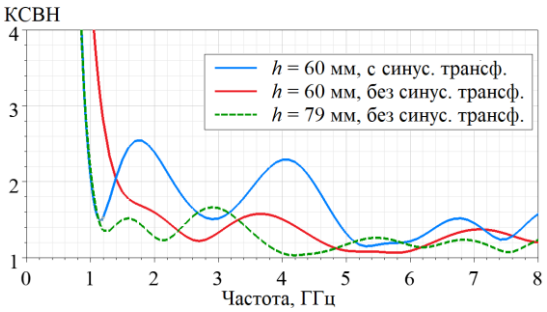



*Рис. 2. КСВН на входе излучателей Вивальди трёх элементарных ячеек*

Для элементарной ячейки высотой 60 мм с синусоидальными симметрирующими трансформаторами нижняя граница рабочей полосы частот, определяемая по уровню КСВН=3, находится на частоте 915 МГц. Для элементарной ячейки такой же высоты с традиционными симметрирующими трансформаторами (без синусоиды) нижняя граница определяется на частоте 1170 МГц (см. рис. 2).

Чтобы получить нижнюю границу рабочей полосы частот АР на частоте 915 МГц с традиционными излучателями Вивальди (без синусоиды), их длина должна составлять 79 мм при прочих неизменных геометрических параметрах. Это на 31% больше решения с симметрирующим трансформатором в форме синусоиды.

Зависимость коэффициента полезного действия (КПД) элементарной ячейки двухполяризационной АР Вивальди с уменьшенной высотой профиля от частоты приведена на рис. 3. В диапазоне рабочих частот 915–7500 МГц КПД превышает 75%.

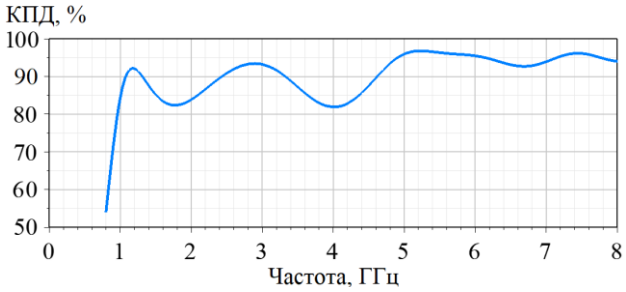



*Рис. 3. КПД элементарной ячейки АР Вивальди с уменьшенной высотой профиля*

Зависимость коэффициента усиления (КУ) элементарной ячейки от частоты дана на рис. 4. Для сравнения приводится теоретический коэффициент направленного действия (КНД) $D$ апертуры такой же площади $A$, что и элементарная ячейка, определяемый по формуле [20]:

$$D = 4\pi A \cos\theta/\lambda^2,$$

где $\theta$ – угол сканирования луча (принят равным нулю) и $\lambda$ – длина волны. Из рис. 4 видно, что КУ элементарной ячейки близок к теоретическому КНД апертуры площади $A$. Понижение КУ наблюдается на частотах, где согласование на входе излучателя Вивальди было наихудшим (см. рис. 2).

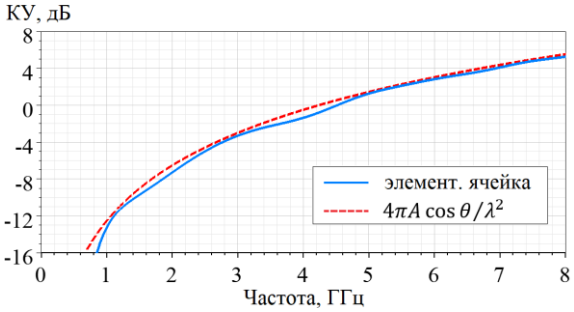



*Рис. 4. КУ элементарной ячейки двухполяризационной АР Вивальди с уменьшенной высотой профиля в сравнении с КНД апертуры такой же площади*

Важно исследовать вопросы согласования и развязки между ортогональными излучателями элементарной ячейки при сканировании луча ДН в пространстве. Ниже даны результаты, полученные для режима возбуждения одного излучателя элементарной ячейки, в то время как второй, ортогональный, излучатель нагружен на согласованную нагрузку.

На рис. 5 показаны зависимости КСВН от частоты при различных углах сканирования ДН в *E*-плоскости. Кривые для углов более 60º не приводятся, поскольку уже при наклоне луча на 60º КСВН на входе излучателя элементарной ячейки превышает 5. Тем не менее, из рис. 5 видно, что эффекта «ослепления» АР не наблюдается во всём секторе углов сканирования.

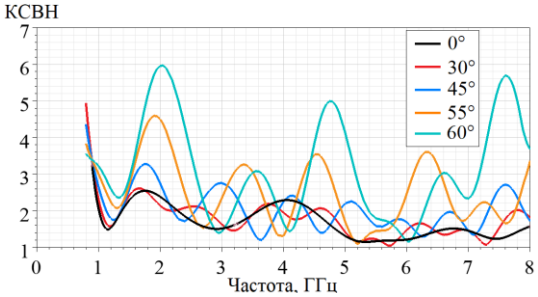



*Рис. 5. КСВН на входе излучателя двухполяризационной АР Вивальди при различных углах сканирования в E-плоскости*

Развязка между ортогональными излучателями элементарной ячейки при различных углах сканирования ДН в *E*-плоскости приводится на рис. 6. На частотах до 1300 МГц развязка составляет от -13 до -20 дБ. На остальных частотах развязка превышает -20 дБ для всех углов сканирования, за исключением локального превышения на частотах 6600-7070 МГц при угле 60º.

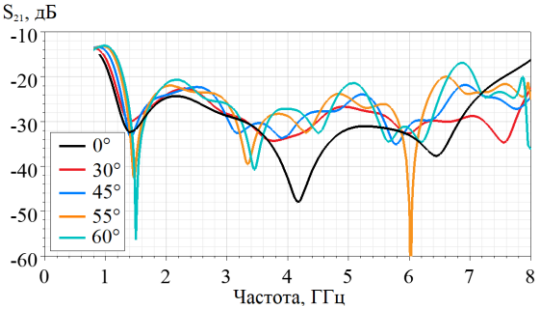



*Рис. 6. Развязка между ортогональными портами элементарной ячейки двухполяризационной АР Вивальди при различных углах сканирования в E-плоскости*

Рассчитанные значения КСВН при сканировании луча в *H*-плоскости показаны на рис. 7. В нижней части рабочего диапазона частот согласование на входе излучателя элементарной ячейки с увеличением угла сканирования ухудшается. Такое сужение рабочей полосы частот при сканировании луча ДН в *H*-плоскости является характерным для многих фазированных АР [21].

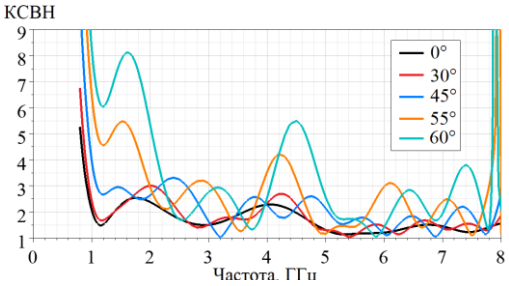



*Рис. 7. КСВН на входе излучателя двухполяризационной АР Вивальди при различных углах сканирования в H-плоскости*

Развязка между ортогональными портами элементарной ячейки при сканировании в *H*-плоскости (рис. 8) качественно повторяет значения, полученные при сканировании луча ДН в *E*-плоскости.

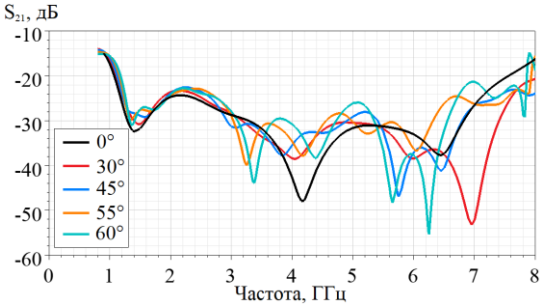



*Рис. 8. Развязка между ортогональными портами элементарной ячейки двухполяризационной АР Вивальди при различных углах сканирования в H-плоскости*

При сканировании луча ДН в *D*-плоскости (диагональная плоскость; угол $\varphi$=45º) эффекта «ослепления» АР не возникает во всем исследуемом секторе углов (рис. 9). Согласование на входе излучателя элементарной ячейки оказывается лучше, чем при сканировании в *H*-плоскости.

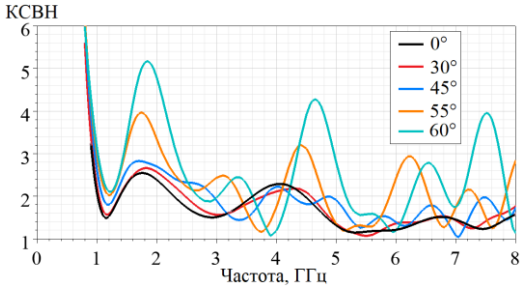



*Рис. 9. КСВН на входе излучателя двухполяризационной АР Вивальди при различных углах сканирования в D-плоскости*

Такие результаты могут ввести в заблуждение и создать ложное ощущение хорошей производительности АР при сканировании в диагональной плоскости. Однако из-за малой развязки между излучателями при сканировании в *D*-плоскости (рис. 10) значительная часть подводимой к излучателю мощности не излучается в свободное пространство, а поступает на соседние излучатели АР [21].

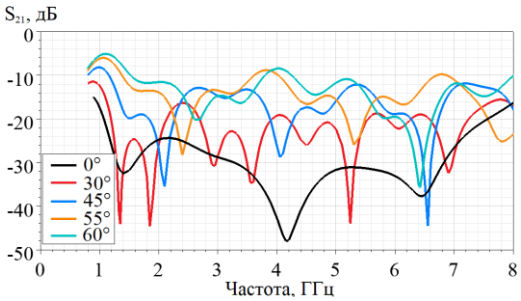



*Рис. 10. Развязка между ортогональными портами элементарной ячейки двухполяризационной АР Вивальди при различных углах сканирования в D-плоскости*

В качестве примера можно рассмотреть частоту 1.2 ГГц при угле сканирования 60º. КСВН на этой частоте равен 2.1 (см. рис. 9), что соответствует 0.58 дБ потерь на рассогласование. Однако развязка между ортогональными портами элементарной ячейки составляет всего минус 5 дБ (см. рис. 10). По этой причине из зависимостей КПД от частоты при различных углах сканирования в D-плоскости (рис. 11) видно, что КПД излучателя элементарной ячейки на частоте 1.2 ГГц составляет только 55%.

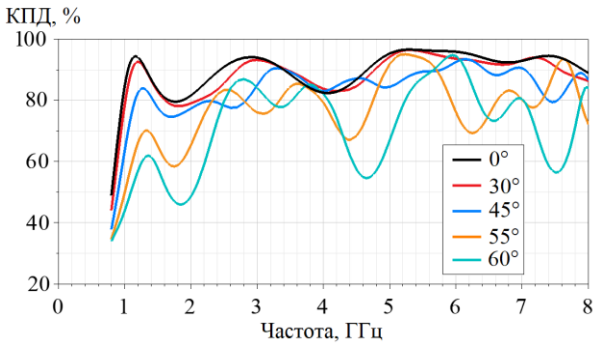



*Рис. 11. КПД излучателя элементарной ячейки двухполяризационной АР Вивальди при различных углах наклона луча ДН в D-плоскости*

АР Вивальди из-за большой электрической длины излучателей отличаются высоким уровнем кроссполяризации. Для АР Вивальди с перекрытием полосы частот более 7:1 характерно превышение уровня кросс-поляризации над уровнем ко-поляризации при сканировании в *D*-плоскости [5]. Однако за счёт уменьшенной высоты профиля и применения симметрирующего трансформатора в форме синусоиды, при наклоне луча на 45º в *D*-плоскости, КУ элементарной ячейки на кросс-поляризации оказывается меньше КУ на ко-поляризации на значение от 6 до 15 дБ (рис. 12).

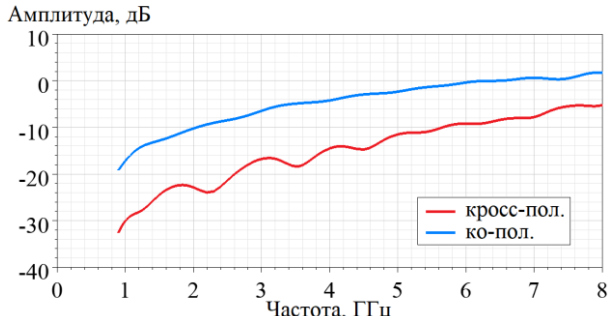



*Рис. 12. КУ на ко- и кросс-поляризации элементарной ячейки двухполяризационной АР Вивальди при наклоне луча на 45º в D-плоскости*

**Выводы.** Исследована элементарная ячейка двухполяризационной АР Вивальди, в которой для уменьшения продольных размеров излучателей применены распределенные симметрирующие трансформаторы в форме синусоиды.

Полоса рабочих частот, определяемая по уровню КСВН≤3 составляет от 915 до 7500 МГц. При этой ширине полосы рабочих частот элементарная ячейка имеет на 31% меньшую высоту, чем элементарная ячейка конвенциональной АР антиподальных излучателей Вивальди. Коэффициент усиления элементарной ячейки принимает значения, близкие к КНД апертуры такой же площади, при этом КПД в полосе рабочих частот превышает 75%.

Результаты расчётов показывают отсутствие эффекта «ослепления» АР при сканировании ДН в *E*-, *H*- и *D*-плоскостях вплоть до угла 60º.

Показано, что, несмотря на хорошее согласование, малая развязка между ортогональными портами элементарной ячейки обеспечивает снижение КПД при сканировании в D-плоскости.

*Исследование выполнено за счёт гранта Российского научного фонда №22-19-00537, https://rscf.ru/project/22-19-00537/ в Центре коллективного пользования «Прикладная электродинамика и антенные измерения Южного Федерального Университета».*

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Gibson P.J.* The Vivaldi Aerial // 1979 9th European Microwave Conf. Brighton, UK, 17–20 Sept. 1979. – IEEE, 1979. – P. 101-105.
2. *Joon Shin, Schaubert D.H.* A parameter study of stripline-fed Vivaldi notch antenna arrays // IEEE Transactions on Antennas and Propagation. – 1999. – Vol. 47, No. 5. – P. 879-886.
3. *Ruiter M., van der Wal E.* EMBRACE, a 10000 element next generation aperture array telescope // 2009 European Microwave Conference (EuMC), Rome, Italy, 2009. – P. 326-329.
4. *McGrath D.T., Schuneman N., Shively T.H., Irion J.* Polarization properties of scanning arrays // IEEE Intern. Symp. on Phased Array Systems and Technology, Boston, *USA*, 14–17 Oct. 2003. – IEEE, 2003. – P. 295-299.
5. *Logan J.T., Kindt R.W., Vouvakis M.N.* Low Cross-Polarization Vivaldi Arrays // IEEE Transactions on Antennas and Propagation. – 2018. – Vol. 66, No. 4. – P. 1827-1837.
6. *Logan J.T., Kindt R.W., Vouvakis M.N.* A 1.2–12 GHz Sliced Notch Antenna Array // IEEE Transactions on Antennas and Propagation. – 2018. – Vol. 66, No. 4. – P. 1818-1826.
7. *Bobkov I.N., Sobol D.D., Yukhanov Y.V.* Low Cross-Polarization Vivaldi Antenna Fed with CPW–Slotline Transition // 2024 Conference of Young Researchers in Electrical and Electronic Engineering (ElCon), Saint Petersburg, Russian Federation, 2024. – P. 629-632.
8. *Kindt R.W., Logan J.T.* Dual-Polarized Metal-Flare Sliced Notch Antenna Array // IEEE Transactions on Antennas and Propagation. – 2020. – Vol. 68, No. 4. – P. 2666-2674.
9. *Бобков И.Н., Юханов Ю.В.* Исследование характеристик элемента плоской антенной решетки Вивальди с расширенной полосой рабочих частот // Известия высших учебных заведений России. Радиоэлектроника. – 2024. – 27 (1). – С. 48-56.
10. *Kindt R.W., Logan J.T.* Single-Polarization Vivaldi Antenna Array with Orthogonal Walls for Improved Polarization Purity // IEEE International Symposium on Phased Array Systems & Technology (PAST), Waltham, MA, USA, 2022. – P. 1-4.
11. *Kindt R.W., Logan J.T.* Cross-Polarization Treatment in Linearly Polarized Vivaldi Array Apertures // IEEE International Symposium on Phased Array Systems & Technology (PAST), Waltham, MA, USA, 2022. – P. 01-04.
12. *Guo L., Qiang Y. F.* Design of a Compact Wideband Dual-Polarization Antipodal Vivaldi Antenna Array // 2018 IEEE International Conference on Computational Electromagnetics (ICCEM), Chengdu, China, 2018. – P. 1-3.
13. *Yang Yunqiang, Wang Y., Aly Fathy* Design of Compact Vivaldi Antenna Arrays for UWB See through Wall Applications // Progress In Electromagnetics Research. – 2008. – Vol. 82. – P. 401-418.
14. *Hodgkinson C.J., Anagnostou D.E., Podilchak S.K.* Compact Antipodal Vivaldi Array with UWB Beam Steering and Element AMC Inclusions for Scattering Reduction // in IEEE Access.
15. *Liu, J., Xu C., Yu H. et al.* Design of a miniaturized ultrawideband and low scattering antipodal vivaldi antenna array // Sci Rep. – 2021. – 11. – 12499.
16. *Elsallal M.W., Schaubert D.H.* Electronically scanned arrays of dual-polarized, doubly-mirrored balanced antipodal Vivaldi antennas (DmBAVA) based on modular elements // 2006 IEEE Antennas and Propagation Society International Symposium, Albuquerque, NM, USA, 2006. – P. 887-890.
17. *Gazit E*. Improved design of the Vivaldi antenna // IEE Proc. Microwaves, Antennas and Propagation. – 1988. – Vol. 135, No. 2. – P. 89-92.
18. Патент RU 203479 U1 H01Q 1/38 (2006.01). Модернизированная сверхширокополосная антенна Вивальди / *Юханов Ю.В., Привалова Т.Ю., Мерглодов И.В., Ильин И.В., Бобков И.Н.*; опубл. 07.04.2021.
19. *Yukhanov Y.V., Bobkov I.N.* Linear Vivaldi Antenna Array with Improved Low-Band Performance // 2021 Radiation and Scattering of Electromagnetic Waves (RSEMW), Divnomorskoe, Russia. – 2021. – P. 203-206.
20. *Pozar D.M.* The active element pattern // IEEE Transactions on Antennas and Propagation. – 1994. – Vol. 42, No. 8. – P. 1176-1178.
21. *Kindt R.W., Logan J.T.* Benchmarking Ultrawideband Phased Antenna Arrays: Striving for Clearer and More Informative Reporting Practices // IEEE Antennas and Propagation Magazine. – 2018. – Vol. 60, No. 3. – P. 34-47.

REFERENCES


1. *Gibson P.J.* The Vivaldi Aerial, *1979 9th European Microwave Conf. Brighton, UK, 17–20 Sept. 1979.* IEEE, 1979, pp. 101-105.
2. *Joon Shin, Schaubert D.H.* A parameter study of stripline-fed Vivaldi notch antenna arrays, *IEEE Transactions on Antennas and Propagation,* 1999, Vol. 47, No. 5, pp. 879-886.
3. *Ruiter M., van der Wal E.* EMBRACE, a 10000 element next generation aperture array telescope, *2009 European Microwave Conference (EuMC), Rome, Italy, 2009,* pp. 326-329.
4. *McGrath D.T., Schuneman N., Shively T.H., Irion J.* Polarization properties of scanning arrays, IEEE *Intern. Symp. on Phased Array Systems and Technology, Boston, USA*, 14–17 Oct. 2003. IEEE, 2003, pp. 295-299.
5. *Logan J.T., Kindt R.W., Vouvakis M.N.* Low Cross-Polarization Vivaldi Arrays, *IEEE Transactions on Antennas and Prop*agation, 2018, Vol. 66, No. 4, pp. 1827-1837.
6. *Logan J.T., Kindt R.W., Vouvakis M.N.* A 1.2–12 GHz Sliced Notch Antenna Array, *IEEE Transactions on Antennas and Propagation*, 2018, Vol. 66, No. 4, pp. 1818-1826.
7. *Bobkov I.N., Sobol D.D., Yukhanov Y.V.* Low Cross-Polarization Vivaldi Antenna Fed with CPW–Slotline Transition, *2024 Conference of Young Researchers in Electrical and Electronic Engineering (ElCon), Saint Petersburg, Russian Federation*, 2024, pp. 629-632.
8. *Kindt R.W., Logan J.T.* Dual-Polarized Metal-Flare Sliced Notch Antenna Array, *IEEE Transactions on Antennas and Propagation*, 2020, Vol. 68, No. 4, pp. 2666-2674.
9. *Bobkov I.N., Yukhanov Yu.V.* Issledovanie kharakteristik elementa ploskoy antennoy reshetki Vival'di s rasshirennoy polosoy rabochikh chastot [Study of characteristics of an element of a flat Vivaldi antenna array with an extended operating frequency band], *Izvestiya vysshikh uchebnykh zavedeniy Rossii. Radioelektronika* [News of higher educational institutions of Russia. Radio electronics], 2024, 27 (1), pp. 48-56.
10. *Kindt R.W., Logan J.T.* Single-Polarization Vivaldi Antenna Array with Orthogonal Walls for Improved Polarization Purity, *IEEE International Symposium on Phased Array Systems & Technology (PAST), Waltham, MA, USA*, 2022, pp. 1-4.
11. *Kindt R.W., Logan J.T.* Cross-Polarization Treatment in Linearly Polarized Vivaldi Array Apertures, *IEEE International Symposium on Phased Array Systems & Technology (PAST), Waltham, MA*, *USA,* 2022, pp. 01-04.
12. *Guo L., Qiang Y.F.* Design of a Compact Wideband Dual-Polarization Antipodal Vivaldi Antenna Array, *2018 IEEE International Conference on Computational Electromagnetics (ICCEM), Chengdu, China*, 2018, pp. 1-3.
13. *Yang Yunqiang, Wang Y., Aly Fathy* Design of Compact Vivaldi Antenna Arrays for UWB See through Wall Applications, *Progress In Electromagnetics Research*, 2008, Vol. 82, pp. 401-418,.
14. *Hodgkinson C.J., Anagnostou D.E., Podilchak S.K.* Compact Antipodal Vivaldi Array with UWB Beam Steering and Element AMC Inclusions for Scattering Reduction, *in IEEE Access*.
15. *Liu, J., Xu C., Yu H. et al.* Design of a miniaturized ultrawideband and low scattering antipodal vivaldi antenna array, *Sci Rep*. – 2021. – 11. – 12499.
16. *Elsallal M.W., Schaubert D.H.* Electronically scanned arrays of dual-polarized, doubly-mirrored balanced antipodal Vivaldi antennas (DmBAVA) based on modular elements, *2006 IEEE Antennas and Propagation Society International Symposium, Albuquerque, NM, USA*, 2006, pp. 887-890.
17. *Gazit E*. Improved design of the Vivaldi antenna, *IEE Proc. Microwaves, Antennas and Propagation,* 1988, Vol. 135, No 2, pp. 89-92.
18. Yukhanov Yu.V., Privalova T.Yu., Merglodov I.V., Il'in I.V., Bobkov I.N. Patent RU 203479 U1 H01Q 1/38 (2006.01). Modernizirovannaya sverkhshirokopolosnaya antenna Vival'di [Patent RU 203479 U1 H01Q 1/38 (2006.01). Upgraded ultra-wideband Vivaldi antenna]; publ. 04/07/2021.
19. *Yukhanov Y.V., Bobkov I.N.* Linear Vivaldi Antenna Array with Improved Low-Band Performance, *2021 Radiation and Scattering of Electromagnetic Waves (RSEMW), Divnomorskoe, Russia*, 2021, pp. 203-206.
20. *Pozar D.M.* The active element pattern, *IEEE Transactions on Antennas and Propagation,* 1994. Vol. 42, No. 8, pp. 1176-1178.
21. *Kindt R.W., Logan J.T.* Benchmarking Ultrawideband Phased Antenna Arrays: Striving for Clearer and More Informative Reporting Practices, *IEEE Antennas and Propagation Magazine*, 2018, Vol. 60, No. 3, pp. 34-47.


Статью рекомендовал к опубликованию д.т.н., профессор К.Е. Румянцев.

**Бобков Иван Николаевич** – Южный федеральный университет; e-mail: antennadesign@outlook.com; г. Таганрог, Россия; тел.: +78634371634; кафедра АиРПУ; аспирант.

**Юханов Юрий Владимирович** – e-mail: yu_yukhanov@mail.ru; тел.: +78634371634; кафедра АиРПУ; д.т.н.; профессор; зав. кафедрой.

**Bobkov Ivan Nikolaevich** – Southern Federal University; e-mail: antennadesign@outlook.com; Taganrog, Russia; phone: +78634371634; the Department of Antennas and Radio Transmitting Devices; graduate student.

**Yukhanov Yury Vladimirovich** – e-mail: yu_yukhanov@mail.ru; phone: +78634371634; the Department of Antennas and Radio Transmitting Devices; dr. of eng. sc.; professor; head of the department.

УДК 621.382 DOI 10.18522/2311-3103-2024-5-214-232

**А.А. Жук**

# СХЕМОТЕХНИЧЕСКИЕ МЕТОДЫ ПОВЫШЕНИЯ ЭФФЕКТИВНОСТИ ВЫХОДНЫХ КАСКАДОВ АРСЕНИД-ГАЛЛИЕВЫХ ОПЕРАЦИОННЫХ УСИЛИТЕЛЕЙ НОВОГО ПОКОЛЕНИЯ*

*Разработка и проектирование арсенид-галлиевых (GaAs) аналоговых функциональных узлов в современной микроэлектронике (операционных усилителях, выходных каскадах, и др.) находится на начальном этапе развития. Это связано с тем, что GaAs широкозонные полупроводники в настоящее время позиционируются преимущественно для сильноточной и сверхвысокочастотной электроники (например, применения в источниках питания, усилителях мощностии т.п.). Для создания микромощной аналоговой компонентной базы, работающей в тяжелых условиях эксплуатации, например, при воздействии высоких температур (+300…+350°С) и радиации, необходима разработка специальных GaAs схемотехнических решений, учитывающих параметры и ограничения соответствующих технологических процессов. Предлагается семейство выходных каскадов, защищенных 5 патентами РФ, для различных модификаций GaAs микромощных операционных усилителей, которые могут быть реализованы на совмещенном GaAs технологическом процессе, позволяющем создавать n-канальные полевые транзисторы с управляющим p-n переходом и GaAs биполярные p-n-p транзисторы. Рассматриваемые схемы выходных каскадов отличаются друг от друга величинами входных и выходных сопротивлений, статическим током потребления, схемотехникой цепей установления статического режима, частотным диапазоном, максимальными амплитудами положительного и отрицательного выходного напряжения и т.п. Приведены результаты сравнительного компьютерного моделирования статического режима, амплитудных и амплитудно-частотных характеристик выходных каскадов в среде LTspice. Предлагаемые схемотехнические решения рекомендуются для применения в GaAs микромощных операционных усилителях нового поколения, а также для использования в составе различных GaAs аналоговых микроэлектронных устройств, в т.ч. работающих в тяжелых условиях эксплуатации: воздействия проникающей радиации и низких температур. При мелкосерийном производстве предложенных выходных каскадов рекомендовано их выполнение на GaAs технологическом процессе, осваиваемом Минским Научно-Исследовательским Институтом Радиоматериалов (ОАО «МНИИРМ», г. Минск, Республика Беларусь), который допускает работу предлагаемых схем в условиях высоких температур (до +300…+350 °С), а также при воздействии проникающей радиации с поглощенной дозой гамма-квантов (до 1 Мрад) и потока нейтронов (до $10^{13}$ н/см$^2$).*

*Операционные усилители; выходные каскады; совмещенные технологические процессы; арсенид-галлиевые полевые транзисторы; арсенид-галлиевые биполярные транзисторы.*

* Исследование выполнено за счет гранта Российского научного фонда № 23-79-10069, https://rscf.ru/project/23-79-10069/.

**A.A. Zhuk**

## HIGH-SPEED OUTPUT STAGES OF OPERATIONAL AMPLIFIERS WITH DIFFERENCING CIRCUIT CORRECTION OF TRANSITION PROCESS

*The development and design of gallium arsenide (GaAs) analogue functional units in modern microelectronics (operational amplifiers, output stages, etc.) is at the initial stage of development. This is because GaAs wide-gap semiconductors are currently positioned primarily for high-current and ultra-high-frequency electronics (e.g., power supplies, power amplifiers, etc.). To create micro-power analogue component base operating under severe operating conditions, for example, under high temperatures (+300...+350°C) and radiation, it is necessary to develop special GaAs circuit solutions that take into account the parameters and limitations of the corresponding technological processes. A family of output stages protected by 5 patents of the Russian Federation for various modifications of GaAs micro-power operational amplifiers is proposed, which can be realised on the combined GaAs technological process allowing to create n-channel field-effect transistors with control p-n junction and GaAs bipolar p-n-p transistors. The considered OS circuits differ from each other by the values of input and output resistances, static current consumption, circuitry of static mode establishment circuits, frequency range, maximum amplitudes of positive and negative output voltage, etc. The results of comparative computer modeling of the static mode, amplitude and amplitude-frequency characteristics of the OS in LTspice simulation software are given. The proposed circuit solutions are recommended for application in GaAs micro-power operational amplifiers of new generation, as well as for use in various GaAs analog microelectronic devices, including those operating under severe operating conditions: exposure to penetrating radiation and low temperatures. At small-scale production of the proposed output stages it is recommended to perform them on GaAs technological process mastered by Minsk Scientific Research Institute of Radio Materials (JSC 'MNIIRM', Minsk, Republic of Belarus), which allows the operation of the proposed circuits at high temperatures (up to +300...+350 $^{o}$C), as well as under the influence of penetrating radiation with absorbed dose of gamma-quanta (up to 1 Mrad) and neutron flux (up to $10^{13}$ n/cm$^2$).*

*Operational amplifiers; output stages; combined process technologies; arsenide-gallium field-effect transistors; arsenide-gallium bipolar transistors.*

**Введение.** В настоящее время в российской и зарубежной микроэлектронике уделяется повышенное внимание высокотемпературным арсенид-галлиевым (GaAs) инструментальным и операционным усилителям (ОУ). Данное направление создания электронной компонентной базы для маломощных интегральных микросхем относится к числу наиболее перспективных в различных областях науки и техники: приборостроении (авиационном, аэрокосмическом и др.), автомобильной и электроэнергетической промышленности, добычи полезных ископаемых и т.п. [1, 2].

Для создания аналоговых микросхем в диапазоне температур до 300-350 $^{o}$C [3] и воздействия радиации [4] применяются GaAs транзисторы. В работе [5] приведены примеры построения таких ОУ. Известны публикации, в которых анонсирован GaAs ОУ на биполярных транзисторах для диапазона температур +300$^{o}$C [6]. Одна из особенностей совмещенных GaAs технологических процессов состоит в том, что они накладывают существенные ограничения на типы реализуемых транзисторов и их характеристики. Так, например, GaAs технологический процесс, предлагаемый фирмами США [7–11], а также Минским научно-исследовательским институтом радиоматериалов (ОАО МНИИРМ, г. Минск, https://mniirm.by/) [11], ориентирован на изготовление аналоговых схем, содержащих только полевые GaAs транзисторы с управляющим p-n переходом и биполярные GaAs p-n-p транзисторы. Применение других полупроводниковых приборов не допускается. Это накладывает существенные ограничения на схемотехнику аналоговых устройств, реализуемых по данному технологическому процессу. К числу основных показателей эффективности выходных каскадов ОУ [12] относятся: максимальные токи в нагрузке при заданных напряжениях питания, входные и выходные сопротивления, статический ток потребления, сквозной ток выходных транзисторов, систематическая составляющая напряжения смещения нуля, верхняя граничная частота по уровню 0,7, максимальная скорость нарастания выходного напряжения, максимальные значения выходного напряжения для положительной и отрицательной полярностей входного сигнала и др. [12, 13]. Целью статьи является исследование семейства выходных каскадов (ВК) опера-

ционных усилителей, защищенных 5 патентами РФ, а также анализ результатов их компьютерного моделирования на GaAs технологическом процессе, осваиваемом Минским Научно-Исследовательским Институтом Радиоматериалов (ОАО «МНИИРМ», г. Минск) [11].

**1. Инвертирующий двухтактный выходной каскад на GaAs биполярных транзисторах.** Особенность предлагаемой на рис. 1 схемы двухтактного выходного каскада [14] состоит в том, что он выполняется только на биполярных GaAs транзисторах.

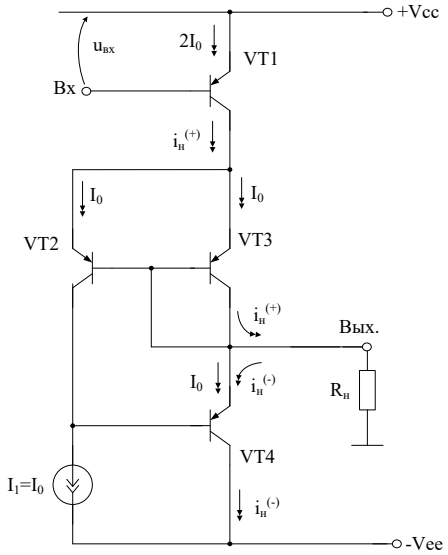



*Рис. 1. Инвертирующий двухтактный выходной каскад на GaAs p-n-p транзисторах [14]*

Схема ВК на рис. 1 содержит входной транзистор VT1, управление которым по цепи базы осуществляется относительно шины положительного источника питания (Vcc). Этот транзистор обеспечивает положительное направление тока $i_н^{(+)}$ в нагрузке Rн. Выходное напряжение и ток $i_н^{(-)}$ отрицательной полярности формируются транзистором VT4.

Статический режим GaAs ВК на рис. 1 определяется двухполюсником $I_1$, который реализуется на JFET транзисторе. При этом в схеме выполняются следующие токовые уравнения Кирхгофа:

$$I_{к2} = I_0 - I_{бп} \approx I_0, \quad (1)$$

$$I_{э3} = I_0, \quad (2)$$

$$I_{э4} = I_0 + I_{бп} \approx I_0, \quad (3)$$

$$I_{э1} = 2I_0. \quad 4)$$

где $I_{бп}$ – ток базы p-n-p транзисторов VT2, VT3, VT4 при токе эмиттера, равном некоторому заданному значению $I_0$, например, 200 мкА.

При положительном приращении входного напряжения на базе VT1 относительно шины питания Vcc через эмиттерно-базовый p-n переход VT3 увеличивается ток $i_н^{(+)}$ в нагрузке $R_н$. Это приводит к увеличению тока эмиттера (коллектора) VT2 и запиранию транзистора VT4. При этом максимальный выходной ток $I_{н.max}^{(+)}$ будет определяться транзистором VT1 и свойствами источника сигнала:

$$I_{н.max}^{(+)} \approx \beta_1 I_{б.max.1}, \quad (5)$$

где $\beta_1$ – коэффициент усиления по току базы VT1; $I_{б.max.1}$ – максимальный ток базы VT1, который определяется свойствами предыдущего каскада усиления.

При отрицательном приращении входного напряжения относительно положительной шины питания – уменьшается коллекторный ток транзистора VT1, а также ток эмиттера (коллектора) биполярных транзисторов VT2 и VT3. В результате VT4 обеспечивает отрицательное приращение тока в нагрузке, причем максимальные значения этого тока:

$$I_{н.max}^{(-)} \approx \beta_4 I_1 = \beta_4 I_0, \quad (6)$$

где $\beta_4$ – коэффициент усиления по току базы транзистора VT4.

На рис. 2 показана схема для моделирования ВК на рис. 1, включённого в структуре GaAs ОУ в среде LTspice.

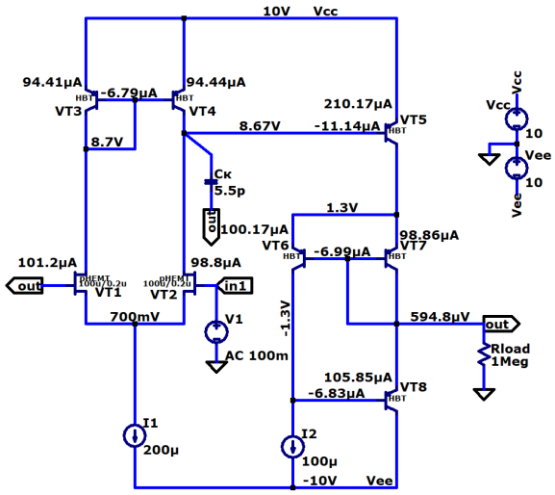



*Рис. 2. Статический режим GaAs ВК на рис. 1 в структуре типового двухкаскадного GaAs ОУ при t=27 °C, $I_1$=200 мкА, $I_2$=100 мкА, $R_{load}$=1 МОм, Ск=5.5 пФ и напряжениях питания ±10 В*

На рис. 3 приведена амплитудная характеристика ОУ на рис. 2 при разных сопротивлениях нагрузки $R_{load}$.

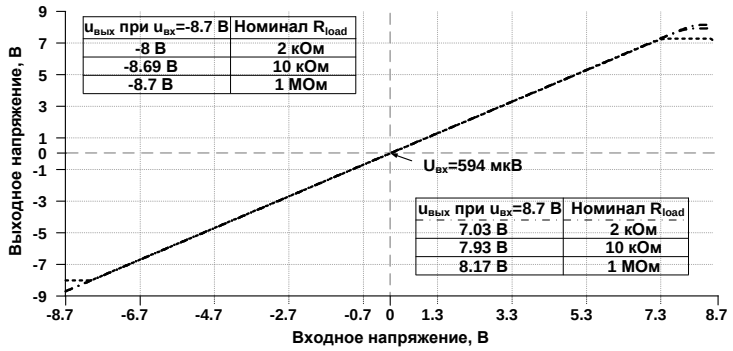



*Рис. 3. Амплитудная характеристика GaAs ОУ с выходным каскадом на рис. 2 в среде LTspice*

На рис. 4 представлена логарифмическая амплитудно–частотная характеристика (ЛАЧХ) коэффициентов усиления по напряжению разомкнутого и замкнутого ОУ (рис. 2). Интегрирующая емкость $C_к$ формирует заданное значение частоты единичного усиления, которая равна $f_1$≈107 МГц.

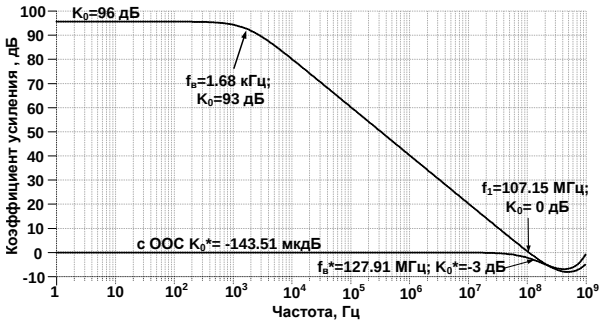



*Рис. 4. ЛАЧХ GaAs ОУ на рис. 2 при t=27 °C*

Компьютерное моделирование схемы GaAs ВК (рис. 2), включенного в состав GaAs ОУ (рис. 3, 4) показывает, что предлагаемая схема [14] обеспечивает двуполярное изменение тока в нагрузке при относительно небольших сопротивлениях $R_н \geq 2$ кОм.

**2. Двухканальный GaAs выходной каскад с высоким входным сопротивлением.** Схема ВК на рис. 5 [14] содержит два канала передачи входного напряжения в нагрузку $R_н$. Для уменьшения входных токов и повышения входного сопротивления в схеме используются GaAs полевые транзисторы VT1 и VT2 с управляющим p-n-переходом.

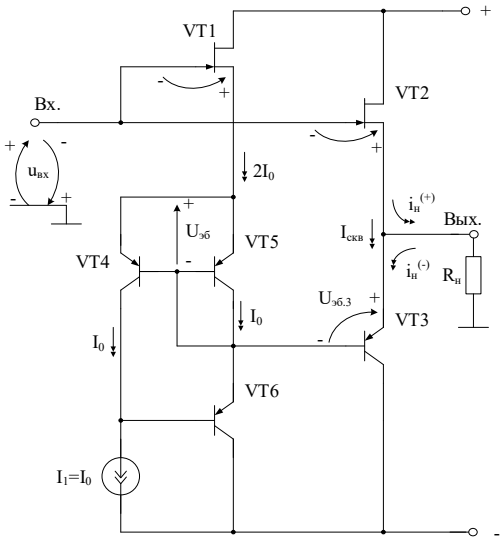



*Рис. 5. GaAs выходной каскад с повышенным входным сопротивлением [14]*

Статический режим двухтактного ВК на рис. 5 определяется двухполюсником $I_1$, который может быть реализован на JFET транзисторе. При этом в схеме выполняются следующие токовые уравнения Кирхгофа:

$$I_{к4} = I_0 - I_{бб} \approx I_0, \quad (7)$$

$$I_{э5} = I_0, \quad (8)$$

$$I_{э3} = I_{и2}, \quad (9)$$

$$I_{и1} = 2I_0 = I_{с1}, \quad (10)$$

$$I_{и2} = I_{э3}. \quad (11)$$

где $I_{и1}=I_{с1}$, $I_{и2}=I_{с2}$– ток истока и ток стока входных полевых транзисторов VT1 и VT2, $I_{бп}$ – ток базы транзисторов VT3, VT4 и VT5 при токе эмиттера, равном заданному значению $I_0$.

При положительном приращении напряжения на входе БУ на рис. 5 увеличивается ток нагрузки двухполюсника $R_н$ через сток полевого транзистора VT2. Как следствие, максимальный положительный ток в нагрузке $I_{н.max}^{(+)}$ будет зависить от ширины каналов полевого транзистора VT2 и рассчитывается по формуле:

$$I_{н.max}^{(+)} = I_{c2.max}, \quad (12)$$

где $I_{c2.max}$– максимальный ток стока полевого транзистора VT2.

При отрицательном приращении входного напряжения БУ на рис. 5 ток в нагрузке $R_н$ будет определяться по формуле, представленной ниже:

$$I_{н.max}^{(-)} = \beta_3 \beta_6 I_0, \quad (13)$$

где $\beta_3, \beta_6$ – коэффициенты усиления по току базы транзисторов VT3 и VT6, $I_0$ – статический ток токостабилизирующего двухполюсника $I_1$.

На рис. 6 показан статический режим ВК рис. 5 в среде LTspice при $I_1$=100 мкА, напряжениях питания ±10 В и высокоомной нагрузке ($R_{load}$ =1 МОм).

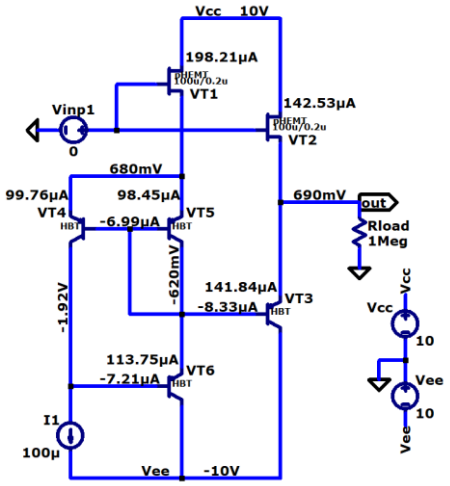



*Рис. 6. Схема для моделирования GaAs ВК на рис. 5 в среде LTspice при t=27°C*

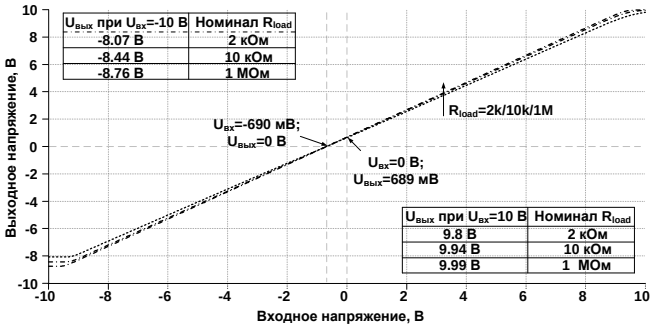



*Рис. 7. Амплитудная характеристика GaAs ВК на рис. 6 в среде LTspice при t=27°C*

Амплитудная характеристика, представленная на рис. 7, показывает, что рассматриваемый ВК обеспечивает выходные напряжения с максимальной амплитудой от -8 В до +9,8 В при $R_{load}$=2 кОм. Для более низкоомных сопротивлений нагрузки необходимо увеличивать ширину канала применяемых полевых транзисторов или использовать параллельное включение нескольких JFET.

**3. GaAs выходной каскад на основе JFET-BJT составного транзистора.** Выходной каскад на рис. 8 [15] содержит биполярные транзисторы VT3, VT4,а также два полевых транзистора VT1, VT2.

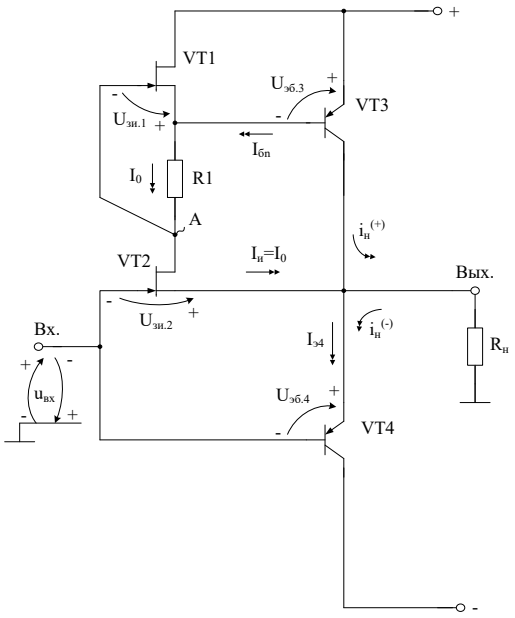



*Рис. 8. GaAs выходной каскад на основе JFET-BJT составного транзистора [15]*

Особенность схемы ВК на рис. 8 состоит в том, что благодаря отрицательной обратной связи по цепи «выход устройства – исток полевого транзистора VT2 – сток полевого транзистора VT2 – затвор полевого транзистора VT1 – база биполярного транзистора VT3 – коллектор биполярного транзистора VT3» здесь устанавливается ток стока VT2, который определяется сопротивлением резистора нагрузки $R_н$:

$$I_{с2} = I_0 = \frac{U_{зи.1}}{R_1}, \quad (14)$$

где $U_{зи.1}$ – напряжение затвор-исток транзистора VT1 при токе истока, равном заданному значению $I_0$.

При высокоомной нагрузке $R_н$ ток эмиттера VT4 имеет две составляющие):

$$I_{э4} = I_{и2} + I_{к3}, \quad (15)$$

где $I_{к3}$ – ток коллектора VT3, $I_{и2}=I_0$ – ток истока VT2.

За счет рационального выбора ширины канала транзистора VT2 и геометрических параметров GaAs биполярного транзистора VT4 можно обеспечить малые (микроамперные) значения тока коллектора $I_{к3}$. В этом случае статический ток потребления схемы на рис. 8 будет определяться сопротивлением R1 и может измеряться десятками микроампер (рис. 9).

Максимальные значения $I_{н.max}^{(+)}$ и $I_{н.max}^{(-)}$ определяются по формулам (16) и (17) соответсвенно:

$$I_{н.max}^{(+)} = I_{R1}\beta_3, \quad (16)$$

$$I_{н.max}^{(-)} = \beta_4 I_{max.ис.с.}, \quad (17)$$

где $I_{R1}$ – ток в согласующем резисторе R1, $\beta_3$, $\beta_4$ – коэффициенты усиления по току базы транзисторов VT3 и VT4, $I_{max.ис.с.}$ – максимальный ток базы VT4, который задается источником входного сигнала.

Статический режим схемы ВК на рис.8 представлен в [15] при параметрах t=27 °C, +Vcc=-Vee=10 В, R1=10 кОм, $R_{load}$=1 МОм.

На рис. 9 приведена амплитудная характеристика GaAs ВК на рис. 8 при разных сопротивлениях нагрузки $R_{load}$=2 кОм/10 кОм/1 МОм.

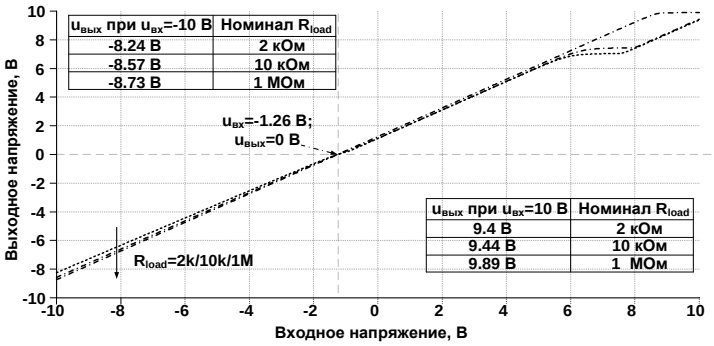



*Рис. 9. Амплитудная характеристика GaAs ВК на рис. 8 при R1=10 кОм при t=27 °C, R1=10 кОм, $R_{load}$=2 кОм/10 кОм/1 МОм и JFET транзисторах с шириной и длиной канала 100u/0.2u*

Биполярный транзистор VT3 в ВК на рис. 8 может быть реализован по схеме составного транзистора Дарлингтона. Такое включение позволяет увеличить значения максимально возможного положительного тока в $R_н$. Дополнительным способом увеличение предельных значений $I^{(+)}_{н.max}$ является уменьшение сопротивления резистора R1, что положительно влияет на амплитудную характеристику, представленную на рис. 10.

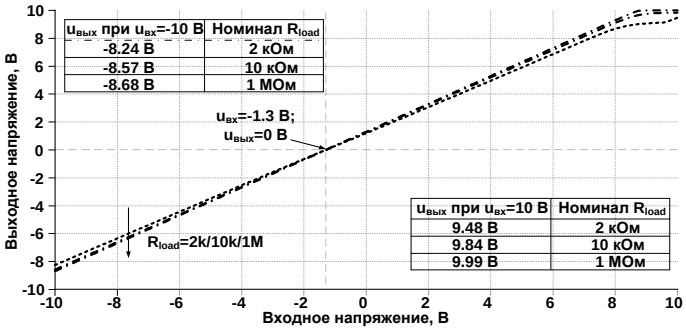



*Рис. 10. Амплитудная характеристика GaAs ВК на рис. 8 при R1=100 Ом*

Логарифмическая амплитудно-частотная характеристика (ЛАЧХ) GaAs ВК на рис. 8 в [15] показывает, что коэффициент передачи по напряжению схемы незначительно отличается от единицы в диапазоне частот до единиц-десятков гигагерц.

Результаты компьютерного моделирования подтверждают то, что рассматриваемая схема ВК [15] обеспечивает двуполярное изменение тока в нагрузке при относительно малых статических токах ее активных элементов и характеризуется коэффициентом передачи по напряжению, близким к единице в широком диапазоне частот. При этом величина сопротивления R1 влияет на максимальные значения выходного напряжения ВК при положительных и отрицательных входных напряжениях.

**4. GaAs выходной каскад с нелинейной коррекцией I-класса.** В основу структуры предлагаемого на рис. 11 GaAs выходного каскада [16] положены идеи нелинейной коррекции [17,18].

При увеличении отрицательного напряжения на входе ВК (рис. 11) подзапирается входной полевой транзистор VT1. Это приводит к увеличению напряжения на стоке VT1 и затворе VT2, а также к увеличению эмиттерного тока биполярного транзистора VT4.

Как следствие, это вызывает увеличение напряжения на затворе выходного полевого транзистора VT3, который «открывается» и создает ток $i_н^{(-)}$ в нагрузке $R_н$. В связи с высоким входным сопротивлением по цепи затвора VT3 ток источника опорного тока $I_1$ может измеряться десятками микроампер.

Если на вход ВК подается положительное напряжение, то это вызывает увеличение тока истока и тока стока транзистора VT1, а также создает ток положительного направления $i_н^{(+)}$ в нагрузке $R_н$. При этом напряжение на стоке VT1 уменьшается, что вызывает запирание VT4 и диода VD2 в режим прямого смещения. При этом $I_{н.max}^{(+)}$ и $I_{н.max}^{(-)}$ определяются по формулам (18) и (19) соответсвенно:

$$I_{н.max}^{(+)} = I_{с1.max}, \tag{18}$$

$$I_{н.max}^{(-)} = \frac{1}{S_3} + r_{VD1}, \tag{19}$$

где $I_{с1.max}$ – максимальный ток стока полевого транзистора VT1, $S_3$ – крутизна стокозатворной характеристкики VT3, $r_{VD1}$– суммарное сопротивление последовательно включеных диодов VD1.

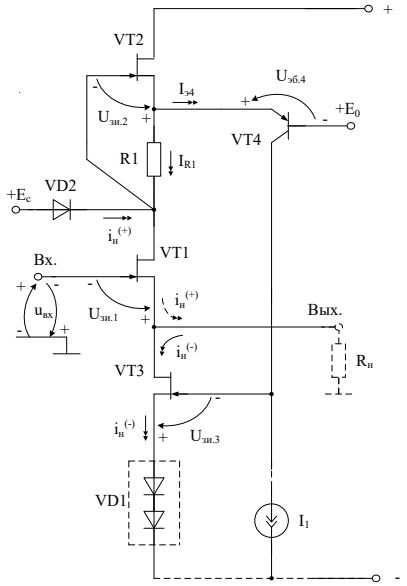



*Рис. 11. GaAs с нелинейной коррекцией I-класса [16]*

Особенность схемы ВК на рис. 12 [16] состоит в том, что здесь источник напряжения смещения $+E_0$ «следит» за уровнем напряжения на выходе ВК. При этом ток положительного направления $i_н^{(+)}$ в нагрузке $R_н$ и ток через p-n переход VD2 замыкаются на низкоомный выход повторителя напряжения (ПН).

Значения $I_{н.max}^{(+)}$ и $I_{н.max}^{(-)}$ ВК на рис. 12 определяются по формулам (20) и (21):

$$I_{н.max}^{(+)} \approx I_{с1.max}, \tag{20}$$

$$I_{н.max}^{(-)} = \frac{u_3}{\frac{1}{S_3} + r_{VD1}} \leq I_{с.max VT3}, \tag{21}$$

где $I_{с1.max}$ – максимальный ток стока полевого транзистора VT1; $u_3$ – напряжение между узлом $\sum \mathbf{I}$ и отрицательной шиной питания, который определяется как: $u_3 = R_i \Delta I$ при $\Delta I = I_0$, где $I_0$ – ток, задающийся транзистором VT2, $S_3$ – крутизна стокозатворной характеристкики VT3; $r_{VD1}$ – суммарное сопротивление диодов VD1; $I_{с.max VT3}$ – максимальный ток стока полевого транзистора VT3.

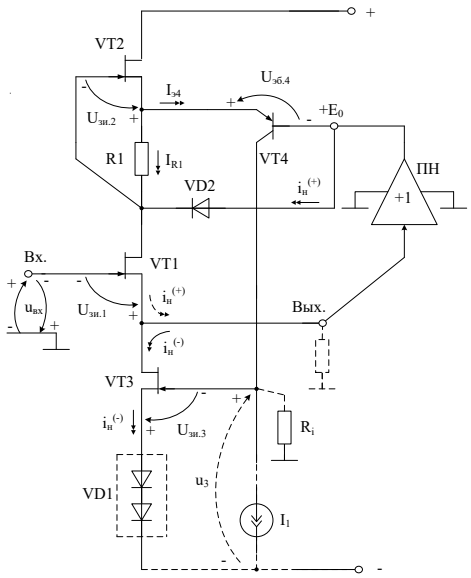



*Рис. 12. GaAs ВК на полевых и биполярных транзисторах со «следящим» источником напряжения смещения $+E_0$ [16]*

В частном случае ПН может быть выполнен на полевых транзисторах, как это сделано на рис. 13 [16], а источник опорного тока $I_0$ (ИОТ) целесообразно выполнять на полевом транзисторе VT4 и резисторе R2.

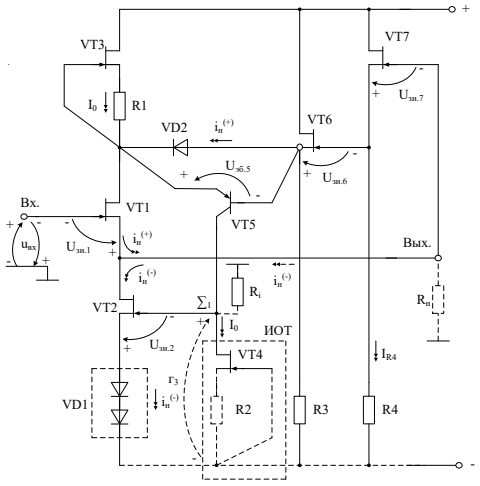



*Рис. 13. Частный случай реализации ПН на n-канальных полевых транзисторах VT6, VT7 [16]*

Величины $I_{н.max}^{(+)}$ и $I_{н.max}^{(-)}$ ВК на рис. 13 определяются по формулам (22) и (23):

$$I_{н.max}^{(+)} \approx I_{с1.max} \approx I_{с6.max}, \quad (22)$$

$$I_{н.max}^{(-)} = \frac{u_3}{\frac{1}{S_2} + r_{VD1}} \leq I_{с.max VT3}, \quad (23)$$

где $I_{с6.max}$ – максимальный ток стока полевого транзистора VT6. В данной схеме ПН на рис. 13 значение $R_i$ в формуле $u_3 = R_i \Delta I$, определяется как (24):

$$R_i = R2 + \frac{1}{S_4}\mu_{VT4}^{-1}, \quad (24)$$

где R2 – значение сопротивления резистора R2, $S_4$ – крутизна стокозатворной характеристики VT4, $\mu_{VT4}^{-1}$ – коэффициент внутренней обратной связи транзистора VT4 в схеме с общим затвором.

Статический режим схемы GaAs ВК на рис. 13 представлен в [16] при параметрах t=27°C, +Vcc=-Vee=10 В, $R_{load}$=1 МОм, R1=R2=11 кОм, R3=R4= 200 кОм, V1=1.8 В.

Амплитудная характеристика ВК на рис. 13 в среде LTspice, представленная на рис. 14, показывает, что рассматриваемая схема при двуполярном питании обеспечивает выходные напряжения с максимальной амплитудой 8,2-9,9 В. Для более низкоомных сопротивлений нагрузки необходимо увеличивать ширину канала применяемых полевых транзисторов VT1, VT2 или использовать параллельное включение нескольких активных JFET.

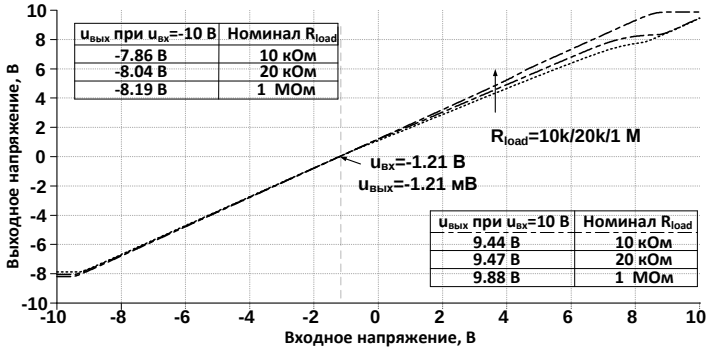



*Рис. 14. Амплитудная характеристика GaAs ВК на рис. 13 в среде LTspice при t=27 °C, V1=1.8 В и разных сопротивлениях нагрузки $R_{load}$=10 кОм/20 кОм/1 МОм*

Компьютерное моделирование на рис. 14 показывает, что предлагаемый ВК, схемотехника которого адаптирована на применение в диапазоне низких температур и воздействия проникающей радиации [3,4], имеет существенные достоинства в сравнении с известными вариантами построения ВК при их реализации в рамках рассматриваемого GaAs технологического процесса.

**5. Двухтактный GaAs выходной каскад с параллельным включением управляющих p-n переходов полевого и биполярного транзисторов.** Особенность схемы на рис. 15 [19] состоит в установлении начального статического тока полевого транзистора VT1 с помощью различных модификаций токостабилизирующих двухполюсников $I_0$ [19]. В схеме на рис. 15 этот токостабилизирующий двухполюсник реализован на транзисторе VT3.

Статический режим ВК на рис. 15 [19] определяется вольт-амперными характеристиками входного полевого транзистора VT1 и биполярного p-n-p транзистора VT2, имеющих различную физическую природу и разные принципы действия, а также GaAs полевого транзистора VT3, формирующего заданный статический ток $I_0$. Причем от свойств этого источника опорного тока $I_0$, который имеет несколько вариантов построения [19], зависят свойства ВК в диапазоне температур и радиационных воздействий [3, 4], а также чувствительность схемы к нестабильности напряжений питания. Значения $I_{н.max}^{(+)}$ и $I_{н.max}^{(-)}$ ВК на рис. 15 определяются по формулам (25) и (26):

$$I_{н.max}^{(+)} = I_{c1.max}, \quad (25)$$

$$I_{н.max}^{(-)} = \beta_2 I_{max.ис.с.}, \quad (26)$$

где $I_{c1.max}$ – максимальный ток стока полевого транзистора VT1, $\beta_2$ – коэффициент усиления по току базы биполярного транзистора VT2, $I_{max.ис.с.}$ – максимальный ток базы VT2, который задается источником входного напряжения.

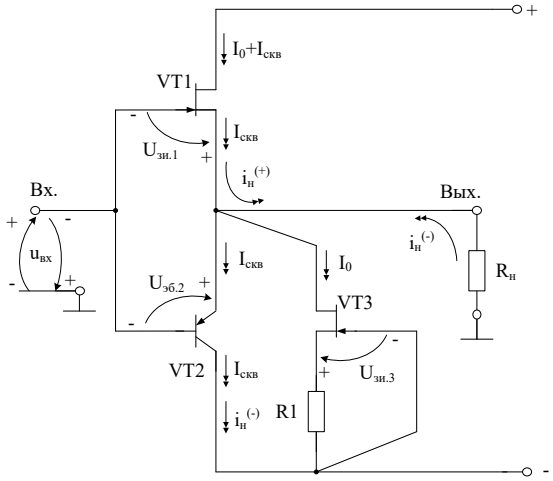



*Рис. 15. Схема GaAs ВК с параллельным включением управляющих p-n переходов JFET и BJT*

Экспериментальные исследования GaAs полевых и биполярных транзисторов показывают, что для получения высокой стабильности сквозного тока ВК ($I_{скв}^{*}$) необходимо иметь вполне определенную температурную или радиационную зависимость тока $I_0$, которая определяется соответствующим выбором рабочей точки на вольт-амперных характеристиках полевого транзистора с n-каналом VT3. Так, применение каскодной структуры на VT3.1 и VT3.2 (рис. 16) вместо одиночного VT3 уменьшает паразитную емкость в цепи нагрузки ВК, что обусловлено эффектом собственной компенсации емкости затвор-сток полевого транзистора VT3.2. Значения $I_{н.max}^{(+)}$ и $I_{н.max}^{(-)}$ ВК на рис. 16 также определяются по формулам (25) и (26):

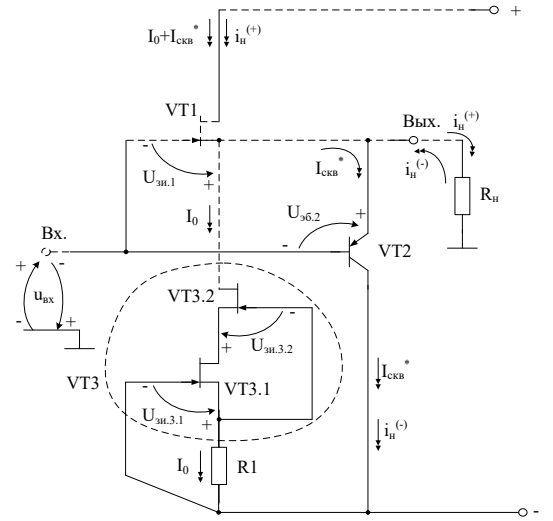



*Рис. 16. GaAs ВК с каскодным включением полевого транзистора VT3 [19]*

Реализация составного split-length транзистора на основе входного VT3.3 и выходного VT3.4 GaAs полевых транзисторов с n-каналом представлена на рис. 17 [19]. В данной модификации можно управлять температурной и радиационной зависимостью статического тока $I_0$ за счет выбора сопротивления резистора R2.

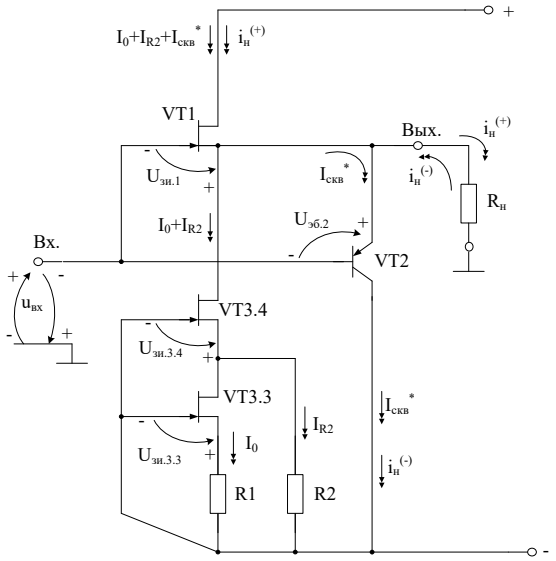



*Рис. 17. Модификация GaAs ВК со split-length составным транзистором VT3.3 и VT3.4 [19]*

Особенность схемы ВК на рис. 18 [19] состоит в том, что за счет применения на входе транзистора VT1 (VT1.1 и VT1.2) в каскодном включении, а также составного транзистора Дарлингтона VT2 (VT2.1 и VT2.2), уменьшается влияние нестабильности напряжения питания на характеристики ВК.

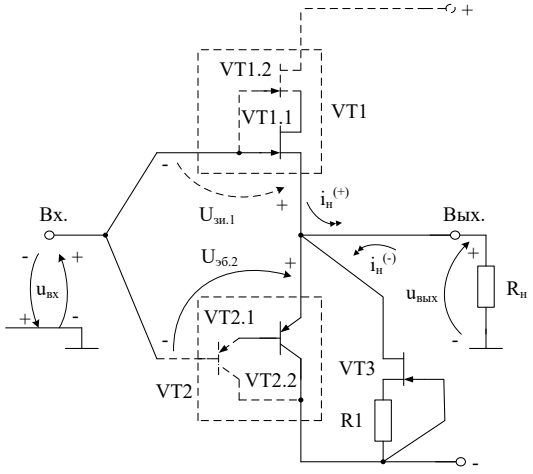



*Рис. 18. Предлагаемый GaAs ВК с split-length составным транзистором VT1 и составным транзистором Дарлингтона (VT21, VT22) [19]*

Если на вход схемы рис. 18 подается положительное напряжение, то это вызывает увеличение тока истока и тока стока VT1, а также создает ток положительного направления $i_н^{(+)}$ в нагрузке $R_н$.

При увеличении отрицательного напряжения на входе ВК подзапирается входной полевой транзистор VT3, а отрицательное приращение тока в нагрузке создается биполярным p-n-p транзистором VT2. Для рассматриваемой схемы ВК на рис. 18 значения $I_{н.max}^{(+)}$ и $I_{н.max}^{(-)}$ определяются формулами (25) и (26).

Статический режим схемы GaAs ВК на рис. 15 представлен в [19] при параметрах t=27 °C, +Vcc=-Vee=10 В, $R_{load}$=1 МОм, R1= 64.5 кОм.

Амплитудная характеристика ВК на рис. 15, представленная на рис. 19, показывает, что рассматриваемая схема при двуполярном питании ±10 В обеспечивает выходные напряжения с максимальной амплитудой 8,2-9,0 В при $R_{load}$≥2 кОм. Для более низкоомных сопротивлений нагрузки $R_{load}$ необходимо увеличивать ширину канала применяемых полевых транзисторов или использовать параллельное включение нескольких JFET, BJT.

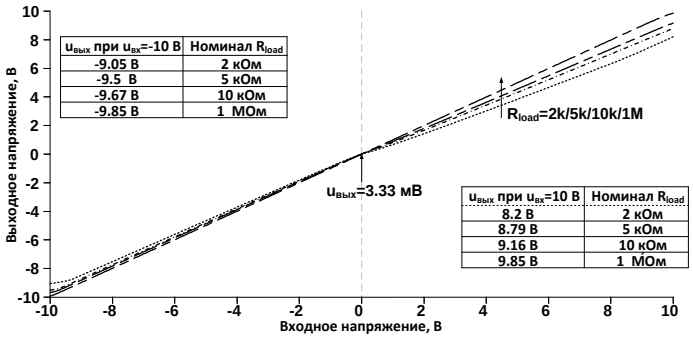



*Рис. 19. Амплитудная характеристика GaAs ВК на рис. 15 в среде LTspice $R_{load}$=2 кОм/5 кОм/10 кОм/1 МОм, V1=-1.24 В, R1= 64.5 кОм*

Из графика амплитудно-частотной характеристики изображенной в [19] следует, что коэффициент передачи по напряжению GaAs ВК на рис. 15 незначительно отличается от единицы в диапазоне частот от единицы до 1 ГГц.

Компьютерное моделирование на рис. 19 демонстрирует, что предлагаемый ВК [19], схемотехника которого адаптирована на применение в диапазоне низких температур и воздействия проникающей радиации [3, 4], имеет существенные достоинства в сравнении с известными вариантами построения ВК при их реализации в рамках рассматриваемого арсенид-галлиевого технологического процесса, обеспечивающего создание только полевых транзисторов с управляющим p-n переходом и биполярных p-n-p транзисторов.

**6. Выходной каскад с входным GaAs биполярным транзистором.** В схеме GaAs ВК на рис. 20 [20] обеспечивается контроль за режимом отсечки выходного биполярного транзистора VT1, что позволяет за счет управления по цепи базы транзистором VT2 обеспечить положительный ток в нагрузке $R_н$.

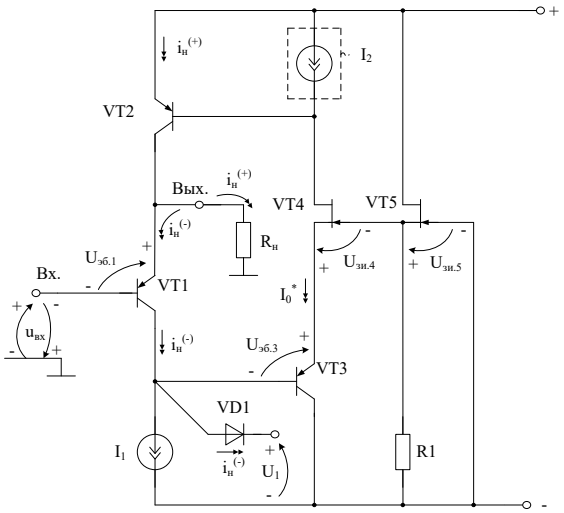



*Рис. 20. GaAs ВК с входным биполярным транзистором VT1 [20]*

За счет отрицательной обратной связи по цепи «биполярный транзистор VT3 – полевой транзистор VT4 – выходной биполярный транзистор VT2» статический ток эмиттера транзистора VT1 определяется током токостабилизирующего двухполюсника $I_1$ и устанавливается на заданном уровне $I_0=I_1$ .

Если на вход подается положительное напряжение относительно общей шины, то это вызывает увеличение тока эмиттера выходного p-n-p транзистора VT2 и формирование положительного тока $i_н^{(+)}$ в нагрузке $R_н$. Максимальное значение тока $I_{н.max}^{(+)}$ зависит от β транзистора VT2, который может выполняться как составной транзистор Дарлингтона (рис. 21).

Когда на базу входного транзистора VT1 подается отрицательное напряжение $u_{вх}^{(-)}$, то этот транзистор формирует отрицательное приращение $i_н^{(-)}$ в нагрузке $R_н$. В этом режиме VT2 запирается и не влияет на работу схемы. Увеличение тока коллектора транзистора VT1 на величину $i_н^{(-)}$ приводит к отпиранию p-n прехода VD1 и, как следствие, большие значения тока $i_н^{(-)} \geq I_0=I_1$ будут «закорачиваться» на вспомогательный источник напряжения $U_1$. Это предотвращает насыщение входного транзистора VT1.

Значения $I_{н.max}^{(+)}$ и $I_{н.max}^{(-)}$ ВК на рис. 20 определяются по формулам (27) и (28):

$$I_{н.max}^{(+)} = \beta_2, \tag{27}$$

$$I_{н.max}^{(-)} = \beta_1 I_{max.ис.с.}, \tag{28}$$

где $\beta_1, \beta_2$ – коэффициенты усиления по току базы биполярных транзисторов VT1 и VT2, $I_{max.ис.с.}$ – максимальный ток базы VT1, который определяется источником входного напряжения.

Для получения повышенных значений $I_{н.max}^{(-)}$ в качестве входного p-n-p транзистора VT1 может применяться составной транзистор по схеме Дарлингтона (рис. 21).

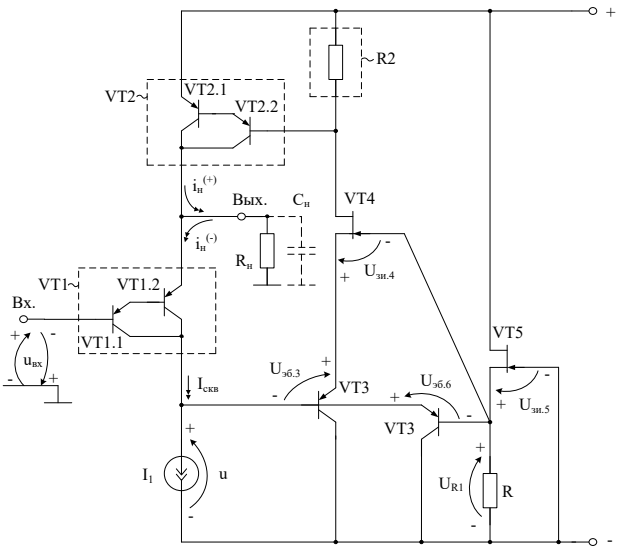



*Рис. 21. Модификация предлагаемой схемы GaAs ВК с составными транзисторами VT1, VT2 [20]*

Значения $I_{н.max}^{(+)}$ и $I_{н.max}^{(-)}$ ВК на рис. 21 также определяются по формулам (29) и (30):

$$I_{н.max}^{(+)} \approx I_1 \beta_3 \beta_2, \tag{29}$$

$$I_{н.max}^{(-)} \approx \beta_1 I_{max.ис.с.}, \tag{30}$$

где $\beta_1, \beta_2, \beta_3$ – коэффициенты усиления по току баз транзисторов VT1÷ VT3; $I_{max.ис.с.}$ – максимальный ток базы VT1, который задается входным источником сигнала.

Амплитудная характеристика ВК на рис. 20, представленная на графиках рис. 22, показывает, что рассматриваемая схема при двуполярном питании ±10В и разных $R_н$ обеспечивает выходные напряжения с максимальной амплитудой от -8,69 В до +9,66 В.

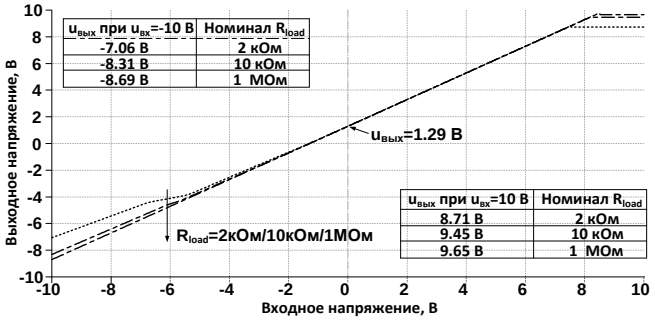



*Рис. 22. Амплитудная характеристика GaAs ВК на рис. 20 в среде LTspice при t=27 °C, $R_н$= 2 кОм/10 кОм/1 МОм, $I_1$= $I_2$=100 мкА, R1=10 кОм, V1=4 В*

На рис. 23 приведена ЛАЧХ GaAs ВК на рис. 20 в среде LTspice при t=27 °C, $R_н$=5 кОм/1 МОм, $I_1$=$I_2$=100 мкА, R1=10 кОм, R2=1 ГОм, V1=4 В. Коэффициент передачи по напряжению предлагаемой схемы ВК на рис. 20 незначительно отличается от единицы в диапазоне частот до 1 ГГц.

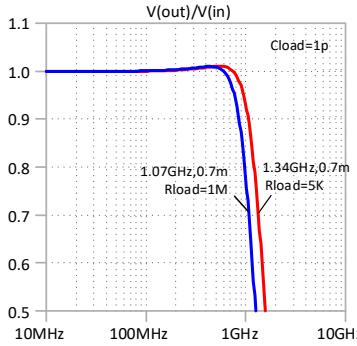



*Рис. 23. ЛАЧХ GaAs ВК на рис. 20*

Статический режим схемы GaAs ВК на рис. 20 представлен в [20] при t=27 °C, +Vcc=-Vee=10 В, $R_н$=1 МОм, $I_1$=100 мкА, R1=14 кОм, R2=10 кОм.

Амплитудная характеристика ВК на рис. 21, представленная на рис. 24, показывает, что рассматриваемая схема при двуполярном питании ±10В и разных $R_н$≥2 кОм обеспечивает выходные напряжения с максимальной амплитудой от -8,69 В до +9,66 В.

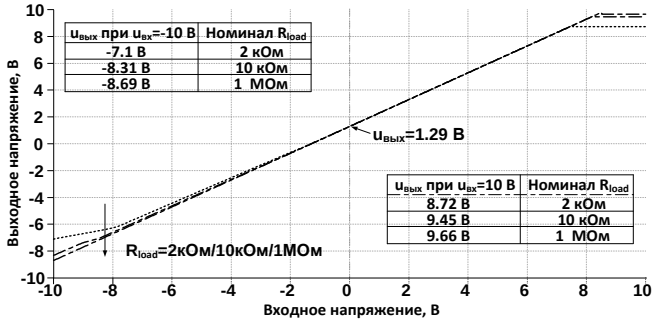



*Рис. 24. Амплитудная характеристика GaAs выходного каскада на рис. 21 в среде LTspice при t=27 °C, $R_н$=2кОм/10кОм/1 МОм, $I_1$=100 мкА, R1=14 кОм, R2=10кОм*

На рис. 25 приведена ЛАЧХ GaAs ВК на рис. 21 в среде LTspice при t= 27°C, $R_н$=5 кОм/1 МОм, $I_1$=100 мкА, R1=14 кОм, R2=10 кОм. Коэффициент передачи по напряжению схемы на рис. 25 также незначительно отличается от единицы в диапазоне частот до 1 ГГц.

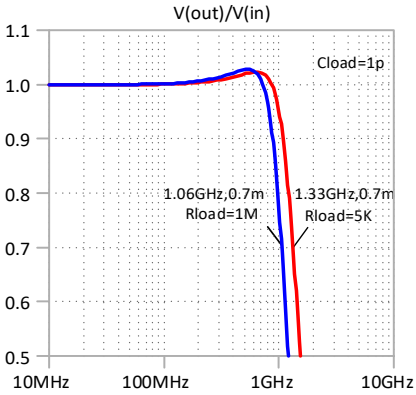



*Рис. 25. Амплитудно-частотная характеристика GaAs выходного каскада на рис. 21 в среде LTspice*

**Заключение.** Рассмотрено семейство GaAs выходных каскадов, предназначеных для использования в структуре маломощных GaAs операционных усилителях нового поколения, а также в различных аналоговых микросхемах, ориентированных на изготовление по совмещенным GaAs технологическим процессам, допускающих формирование на одном кристалле полевых транзисторов с n-каналом и биполярных p-n-p транзисторов. Результаты компьютерного моделирования показывают, что максимальные выходные напряжения предлагаемых схем ВК, которые защищены как объект интеллектуальной собственности 5 патентами России, отличаются от напряжений на соответствующих шинах питания на 10-20%.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Xiao Y. et al.,* Review of High-Temperature Power Electronics Converters // IEEE Transactions on Power Electronics. – 2022. – Vol. 37, No. 12. – P. 14831-14849. – DOI: 10.1109/TPEL.2022.3148192.
2. *Dreike P.L. et al.* An overview of high-temperature electronic device technologies and potential applications // IEEE Transactions on Components, Packaging, and Manufacturing Technology: Part A. – 1994. – Vol. 17, No. 4. – P. 594-609. – DOI: 10.1109/95.335047.
3. *Wiirfl J., Janke B., Nebauer E., Thierbach S. and Wolter P.*, High Temperature MESFET Based Integrated Circuits Operating up to 300°C // IEDM. – P. 96-219.
4. *Song Y., Kim M.E., Oki A.K., Hafizi M.E., Camou J.B. and Kobayashi K.W.* Radiation hardness characteristics of GaAs/AlGaAs heterojunction bipolar transistors // 11th Annual Gallium Arsenide Integrated Circuit (GaAs IC). – 1989. – P. 155-158. – DOI: 10.1109/GAAS.1989.69316.
5. *Beasom J.D., Patterson R.B.* Process Characteristics and Design Methods for a 300°C Quad Operational Amplifier // IEEE Transactions on Industrial Electronics. – 1982. – Vol. 2. – P. 112-117. – DOI: 10.1109/TIE.1982.356646.
6. *Doerbeck F.H., Duncan W.M., Mclevige W.V. and Yuan H.T.* Fabrication and High-Temperature Characteristics of Ion-Implanted GaAs Bipolar Transistors and Ring-Oscillators // IEEE Transactions on Industrial Electronics. – 1982. – Vol. 2. – P. 136-139. – DOI: 10.1109/TIE.1982.356650.
7. *Fresina M.* Trends in GaAs HBTs for wireless and RF // IEEE Bipolar/BiCMOS Circuits and Technology Meeting. – 2011. – P. 150-153. – DOI: 10.1109/BCTM.2011.6082769.
8. *Zampardi P.J., Sun M., Cismaru C., Li J.* Prospects for a BiCFET III-V HBT Process // IEEE Compound Semiconductor Integrated Circuit Symposium (CSICS). – 2012. – P. 1-3. – DOI: 10.1109/CSICS.2012.6340116.
9. *Liu W., Hill D., Costa D., Harris J.S.* High-performance microwave AlGaAs-InGaAs Pnp HBT with high-DC current gain // IEEE Microwave and Guided Wave Letters. – 1992. – Vol. 2, No. 8. – P. 331-333. – DOI: 10.1109/75.153604.

10. *Peatman W. et al.* InGaP-Plus™: advanced GaAs BiFET technology and applications // CS MANTECH Conference. – 2007. – P. 243-246.
11. *Дворников О.В., Павлючик А.А., Прокопенко Н.Н., Чеховский В.А и др.* Унифицированные схемотехнические решения аналоговых арсенид-галлиевых микросхем // Известия вузов. Электроника. – 2022. – Т. 27, № 4. – С. 475-488. – DOI: https://doi.org/10.24151/1561-5405-2022-27-4-475-488.
12. *Carter B., Mancini R.* Op Amps for Everyone. – 2017. – ISBN: 9780128116470.
13. *Эннс В.И., Кобзев Ю.М.* Проектирование аналоговых КМОП-микросхем: краткий справочник разработчика. – 2015. – 2-е изд. – С. 445-446.
14. *Прокопенко Н.Н., Жук А.А., Бугакова А.В.* Арсенид-галлиевый буферный усилитель, № RU 2784046 от 2022.
15. *Прокопенко Н.Н., Жук А.А., Титов А.Е.* Неинвертирующий выходной каскад арсенид-галлиевого операционного усилителя, № RU 2784049 от 2022.
16. *Савченко Е.М., Прокопенко Н.Н., Жук А.А., Пронин А.А. и др.* Арсенид-галлиевый буферный усилитель на полевых и биполярных p-n-p транзисторах, № RU 2788498 от 2023.
17. *Прокопенко Н.Н.* Нелинейная активная коррекция в прецизионных аналоговых микросхемах: монография. – Ростов-на-Дону, 2000. – 223 с.
18. *Прокопенко Н.Н., Никуличев Н.Н.* Нелинейная коррекция на основе управляемых коммутаторов тока и напряжения в аналоговых микросхемах: монография. – Шахты, 2006. – 115 с.
19. *Прокопенко Н.Н., Жук А.А., Кунц А.В., Гавлицкий А.И.* Арсенид-галлиевый буферный усилитель на основе n-канальных полевых и p-n-p биполярных транзисторов, № RU 2784376 от 2022.
20. *Прокопенко Н.Н., Чумаков В.Е., Кунц А.В., Жук А.А.* Арсенид-галлиевый выходной каскад быстродействующего операционного усилителя, № RU 2773912 от 2022.

REFERENCES

1. *Xiao Y. et al.,* Review of High-Temperature Power Electronics Converters, *IEEE Transactions on Power Electronics,* 2022, Vol. 37, No. 12, pp. 14831-14849. DOI: 10.1109/TPEL.2022.3148192.
2. *Dreike P.L. et al.* An overview of high-temperature electronic device technologies and potential applications, *IEEE Transactions on Components, Packaging, and Manufacturing Technology: Part A,* 1994, Vol. 17, No. 4, pp. 594-609. DOI: 10.1109/95.335047.
3. *Wiirfl J., Janke B., Nebauer E., Thierbach S. and Wolter P.*, High Temperature MESFET Based Integrated Circuits Operating up to 300°C, *IEDM,* pp. 96-219.
4. *Song Y., Kim M.E., Oki A.K., Hafizi M.E., Camou J.B. and Kobayashi K.W.* Radiation hardness characteristics of GaAs/AlGaAs heterojunction bipolar transistors, *11th Annual Gallium Arsenide Integrated Circuit (GaAs IC),* 1989, pp. 155-158. DOI: 10.1109/GAAS.1989.69316.
5. *Beasom J.D., Patterson R.B.* Process Characteristics and Design Methods for a 300°C Quad Operational Amplifier, *IEEE Transactions on Industrial Electronics,* 1982, Vol. 2, pp. 112-117. DOI: 10.1109/TIE.1982.356646.
6. *Doerbeck F.H., Duncan W.M., Mclevige W.V. and Yuan H.T.* Fabrication and High-Temperature Characteristics of Ion-Implanted GaAs Bipolar Transistors and Ring-Oscillators, *IEEE Transactions on Industrial Electronics,* 1982, Vol. 2, pp. 136-139. DOI: 10.1109/TIE.1982.356650.
7. *Fresina M.* Trends in GaAs HBTs for wireless and RF, *IEEE Bipolar/BiCMOS Circuits and Technology Meeting,* 2011, pp. 150-153. DOI: 10.1109/BCTM.2011.6082769.
8. *Zampardi P.J., Sun M., Cismaru C., Li J.* Prospects for a BiCFET III-V HBT Process, *IEEE Compound Semiconductor Integrated Circuit Symposium (CSICS),* 2012, pp. 1-3. DOI: 10.1109/CSICS.2012.6340116.
9. *Liu W., Hill D., Costa D., Harris J.S.* High-performance microwave AlGaAs-InGaAs Pnp HBT with high-DC current gain, *IEEE Microwave and Guided Wave Letters,* 1992, Vol. 2, No. 8, pp. 331-333. DOI: 10.1109/75.153604.
10. *Peatman W. et al.* InGaP-Plus™: advanced GaAs BiFET technology and applications, *CS MANTECH Conference,* 2007, pp. 243-246.
11. *Dvornikov O.V., Pavlyuchik A.A., Prokopenko N.N., Chekhovskiy V.A i dr.* Unifitsirovannye skhemotekhnicheskie resheniya analogovykh arsenid-gallievykh mikroskhem [Unified circuit design solutions for analog gallium arsenide microcircuits], *Izvestiya vuzov. Elektronika* [News of universities. Electronics], 2022, Vol. 27, No. 4, pp. 475-488. DOI: https://doi.org/10.24151/1561-5405-2022-27-4-475-488.
12. *Carter B., Mancini R.* Op Amps for Everyoneб 2017. ISBN: 9780128116470.
13. *Enns V.I., Kobzev Yu.M.* Proektirovanie analogovykh KMOP-mikroskhem: kratkiy spravochnik razrabotchika [Design of analog CMOS integrated circuits: a developer's quick reference], 2015. 2nd ed., pp. 445-446.

14. *Prokopenko N.N., Zhuk A.A., Bugakova A.V.* Arsenid-gallievyy bufernyy usilitel', № RU 2784046 ot 2022 [Gallium arsenide buffer amplifier, No. RU 2784046 dated 2022].
15. *Prokopenko N.N., Zhuk A.A., Titov A.E.* Neinvertiruyushchiy vykhodnoy kaskad arsenid-gallievogo operatsionnogo usilitelya, № RU 2784049 ot 2022 [Non-inverting output stage of a gallium arsenide operational amplifier, No. RU 2784049 of 2022].
16. *Savchenko E.M., Prokopenko N.N., Zhuk A.A., Pronin A.A. i dr.* Arsenid-gallievyy bufernyy usilitel' na polevykh i bipolyarnykh p-n-p tranzistorakh, № RU 2788498 ot 2023 [Gallium arsenide buffer amplifier on field-effect and bipolar p-n-p transistors, No. RU 2788498 of 2023].
17. *Prokopenko N.N.* Nelineynaya aktivnaya korrektsiya v pretsizionnykh analogovykh mikroskhemakh: monografiya [Nonlinear active correction in precision analog microcircuits: monograph]. Rostov-on-Don, 2000, 223 p.
18. *Prokopenko N.N., Nikulichev N.N.* Nelineynaya korrektsiya na osnove upravlyaemykh kommutatorov toka i napryazheniya v analogovykh mikroskhemakh: monografiya [Nonlinear correction based on controlled current and voltage switches in analog microcircuits: monograph]. Shakhty, 2006, 115 p.
19. *Prokopenko N.N., Zhuk A.A., Kunts A.V., Gavlitskiy A.I.* Arsenid-gallievyy bufernyy usilitel' na osnove n-kanal'nykh polevykh i p-n-p bipolyarnykh tranzistorov, № RU 2784376 ot 2022 [Gallium arsenide buffer amplifier based on n-channel field-effect and p-n-p bipolar transistors, No. RU 2784376 of 2022].
20. *Prokopenko N.N., Chumakov V.E., Kunts A.V., Zhuk A.A.* Arsenid-gallievyy vykhodnoy kaskad bystrodeystvuyushchego operatsionnogo usilitelya, № RU 2773912 ot 2022 [Gallium arsenide output stage of a high-speed operational amplifier, No. RU 2773912 of 2022].

Статью рекомендовал к опубликованию д.т.н., профессор Б.Г. Коноплев.

**Жук Алексей Андреевич –** Донской государственный технический университет; e-mail: alexey.zhuk96@mail.ru; г. Ростов-на-Дону, Россия; тел.: +79185880301; младший научный сотрудник отдела «Управление научных исследований»; ассистент кафедры «Информационные системы и радиотехника».

**Zhuk Alexey Andreevich –** Don State Technical University; e-mail: alexey.zhuk96@mail.ru; Rostov-on-Don, Russia; phone: +79185880301; junior research Fellow of the “Office of Scientific Research”; assistant of the Department of Information Systems and Radio Engineering.

УДК 621.396.67 DOI 10.18522/2311-3103-2024-5-232-242

**Р.Э. Косак**

## АНТЕННАЯ РЕШЕТКА КОМПАКТНЫХ ИЗЛУЧАТЕЛЕЙ ВИВАЛЬДИ С ЭЛЛИПТИЧЕСКИМИ ВЫРЕЗАМИ НА КРОМКЕ

*Исследовано влияние эллиптических вырезов на кромке компактного излучателя Вивальди, рассчитанного в составе бесконечной фазированной антенной решетки (ФАР) и конечной антенной решетки (АР), на его характеристики излучения. Оценены коэффициент стоячей волны по напряжению (КСВН) и коэффициент усиления (КУ) излучателя. Для излучателя в составе бесконечной ФАР характеристики излучения получены в секторе углов ±60° в плоскостях E и H. Определено, что введение вырезов эллиптической формы размером 3 × 2 мм на кромке излучателя Вивальди в составе бесконечной ФАР позволяет расширить рабочую полосу частот по уровню КСВН ≤ 3 в обеих плоскостях, а также улучшить средний уровень КСВН в E–плоскости. В режиме сканирования в секторе углов ±60° в E–плоскости коэффициент перекрытия по уровню КСВН ≤ 3 увеличивается с 2,86 до 3,41, а в H–плоскости в режиме сканирования в секторе углов ±45° коэффициент перекрытия по уровню КСВН ≤ 3 (≤ 3,05 при 45°) увеличивается с 2,74 до 3,15. Исследована 16-элементная АР компактных сверхширокополосных (СШП) излучателей Вивальди с эллиптическими вырезами на кромке и без них. При добавлении эллиптических вырезов в конструкцию излучателей АР конечного размера коэффициент перекрытия увеличивается с 2,07 до 2,37. Определено, что АР, также как и излучатель в составе бесконечной ФАР, является СШП и может работать в диапазоне от 283,8 до 671,3 МГц по уровню КСВН ≤ 3, чему соответствует коэффициент перекрытия ~ 2,37. Средний уровень КСВН при включении всех излучателей располагается по уровню КСВН = 4, а при подключении одного ряда излучателей на согласованные на-*

*грузки – по уровню КСВН = 1,6. В основном практически во всей рабочей полосе частот в этом случае значение КСВН ниже уровня КСВН = 2,3. КУ в рабочей полосе частот располагается в пределах от 3,82 до 9,50 дБ.*

*Антенная решетка; вырезы эллиптической формы на кромке; компактный излучатель Вивальди; КСВН; коэффициент усиления; режим широкоугольного сканирования; сверхширокая полоса частот.*

**R.E. Kosak**

## ANTENNA ARRAY OF COMPACT VIVALDI RADIATORS WITH ELLIPTICALL SHAPE CUTOUTS ON THEIR OUTER EDGE

*The influence of elliptical cutouts on the outer edge of a compact Vivaldi radiator, designed as part of an infinite phased array (PA) and a finite antenna array (AA), on its radiation characteristics is investigated. The voltage standing wave ratio (VSWR) and realized gain (RG) of the radiator are estimated. For the radiator as part of the infinite PA, the radiation characteristics were obtained in the sector of angles ±60° in the E– and H–planes. The research determined that the introduction of elliptical cutouts measuring 3 × 2 mm on the outer edges of the Vivaldi radiator as part of the PA makes it possible to expand the operating frequency band in terms of VSWR ≤ 3 in both planes, and improve the average level of matching in wide-angle scanning mode in the E-plane. In the E–plane in scanning mode in the sector of angles ±60°, the overlap ratio at the level of VSWR ≤ 3 increases from 2.86 to 3.41, and in the H–plane in scanning mode in the sector of angles ±45°, the overlap ratio at the level of VSWR ≤ 3 (≤ 3,05 at 45°) increases from 2.74 up to 3.15. A 16-element array from the compact ultra-wideband (UWB) Vivaldi radiators with and without elliptical cutouts is researched. When using elliptical cutouts in the design of finite-size antenna array radiators, the overlap ratio increases from 2.07 to 2.37. It has been determined that the AA, as well as the radiator in the PA, is UWB and can operate in the range from 283.8 to 671.3 MHz at the VSWR level ≤ 3, which corresponds to overlap ratio ~ 2.37. The average VSWR level when all radiators are turned on is located at the VSWR level = 4, and when connecting one row of radiators to matched loads – at the VSWR level = 1.6. Basically, over almost the entire operating frequency band in this case, the VSWR value is below the VSWR level = 2.3. The realized gain in the operating frequency band is in the range from 3.82 to 9.50 dB.*

*Antenna array; elliptical cutouts on the outer edge; compact Vivaldi radiator; VSWR; realized gain; wide-angle scanning mode; ultra-wide frequency band.*

**Введение.** В радиолокационных и радионавигационных системах обеспечения аэропортов, системах спутникового вещания и связи с подвижными объектами, системах обеспечения безопасности движения автомобиля, а также радиоастрономических системах широкое применение находят АР и ФАР [1, 2]. Известно применение ФАР в медицине [3, 4], системах цифрового телевидения [5] и спутниковых мобильных телекоммуникационных системах [6].

АР и ФАР, содержащие большое количество излучающих печатных элементов, поддаются анализу на основе теоремы Флоке для периодических ячеек [7]. В качестве излучателей АР и ФАР используют вибраторы, открытые концы волноводов, диэлектрические стержни, спирали и так далее [8]. В последнее время все большее внимание уделяется печатным излучателям, разработанным на основе антенн Вивальди [9–11]. Важную роль играет форма излучателя Вивальди – именно она во многом определяет его характеристики излучения.

Использование вырезов на кромке позволяет расширить рабочую полосу частот, улучшить согласование и КУ [11–19]. Согласно [11], прямоугольная форма вырезов на кромке излучателя Вивальди позволяет уменьшить его электрический размер, что делает его более компактным. Использование прямоугольных вырезов на кромке излучателя также позволяет улучшить КУ и увеличить согласование излучателя, улучшить направленность [12] и уменьшить уровень боковых лепестков [13]. Синусоидальная форма вырезов также позволяет увеличить КУ и уменьшить коэффициент отражения [14]. Более того, по сравнению с прямоугольной формой вырезов, синусоидальная форма может улучшить КУ в рабочей полосе частот в среднем на 3 дБ и коэффициент полезного действия до 0,9 [18].

Влияние размеров прямоугольной формы гофр на коэффициент отражения и КУ оценено в [19]. Показано, что прямоугольная форма гофр с одинаковым размером, а также с уменьшающимся размером в сторону раскрыва антенны Вивальди в рабочей полосе частот приводит к увеличению КУ и снижению коэффициента отражения, а прямоугольной формы гофр с увеличивающимся размером в сторону раскрыва антенны – наоборот.

Таким образом, из представленных работ видно, что изменение кромки лепестков излучателя Вивальди является довольно распространенным способом улучшения характеристик излучения. При этом, прямоугольная форма гофр является наиболее распространенной, а работ с эллиптическими формами гофр или схожими с ними меньше.

**Цель** работы состоит в улучшении согласования и расширении рабочей полосы частот компактного СШП излучателя Вивальди путем введения эллиптических вырезов на его кромке.

**Излучатель Вивальди без эллиптических вырезов на кромке.** В работе за основу взят излучатель Вивальди, представленный на рис. 1 [15]. Подложка выполнена из материала RT/duroid 5880, а питание производится с помощью коаксиального кабеля сопротивлением 50 Ом. Размер ячейки в составе ФАР 100 × 120 × 185 мм. Расчет в составе бесконечной АР с использованием периодических граничных условий на боковых поверхностях ячейки позволяет существенно сократить объем вычислений [20].

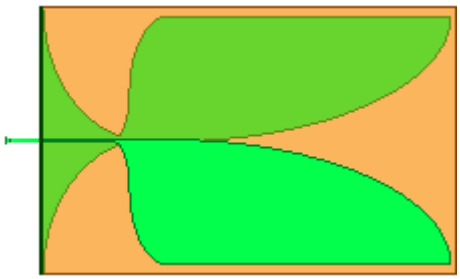
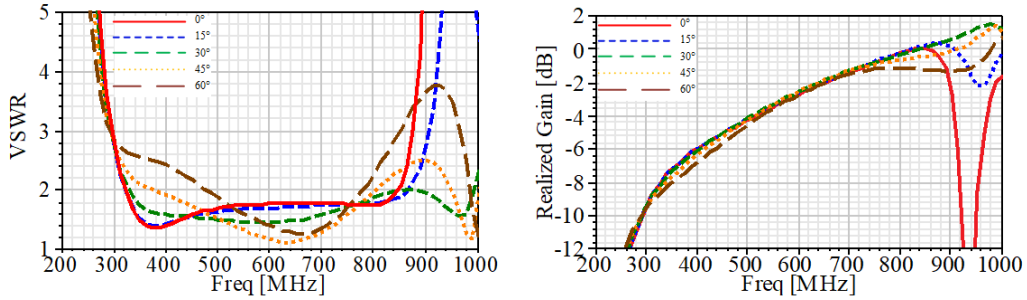


*Рис. 1. Излучатель Вивальди без вырезов на кромке*

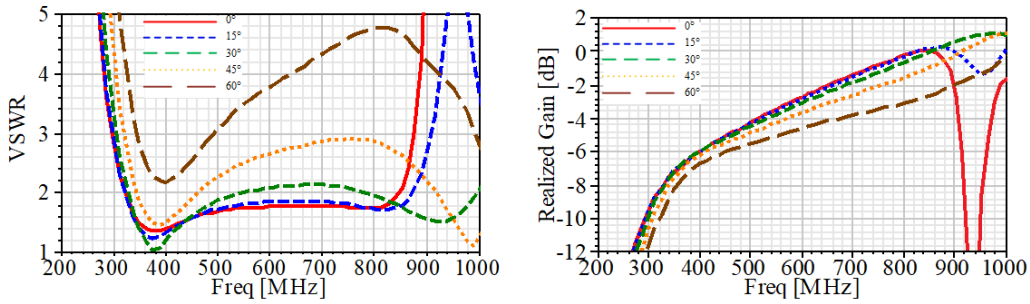
Из [15] следует, что описанная конструкция излучателя Вивальди позволяет добиться сверхширокой рабочей полосы частот в широком секторе углов сканирования в диапазоне ±60°. Также этот излучатель является электрически компактным: его высота на верхней рабочей частоте составляет ~ 0,5 λ.

Рассмотрены КСВН и КУ излучателя в составе бесконечной ФАР в диапазоне от 200 до 1000 МГц в режиме сканирования в плоскостях E и H. На рис. 2, 3 каждой кривой соответствует определённое значение угла сканирования: 0° (—), 15° (---), 30° (– –), 45° (·····) и 60° (– –).




а б

*Рис. 2. КСВН (а) и КУ (б) излучателя Вивальди без эллиптических вырезов на кромке в режиме сканирования в E–плоскости*




а б

*Рис. 3. КСВН (а) и КУ (б) излучателя Вивальди без эллиптических вырезов на кромке в режиме сканирования в Н–плоскости*

Согласно рис. 2,а, рабочая полоса частот излучателя Вивальди (рис. 1) в Е–плоскости находится в пределах от 297 до 848 МГц по уровню КСВН ≤ 3 в режиме сканирования в секторе углов ±60°. Коэффициент перекрытия, определяемый как отношение верхней граничной частоты к нижней, равен 2,86. Как видно из рис. 2,б, КУ в указанной полосе частот находится в пределах от минус 9,71 до 0,28 дБ.

Согласно рис. 3,а, рабочая полоса частот исследуемого излучателя Вивальди в Н–плоскости находится в пределах от 319 до 874 МГц по уровню КСВН ≤ 3 (коэффициент перекрытия ~ 2,74) в режиме сканирования в секторе углов ±45°. КУ в указанной полосе частот, согласно рис. 3,б находится в пределах от минус 9,10 до 0,34 дБ.

**Исследование влияния эллиптических вырезов на кромке излучателя.** Для оценки влияния эллиптических вырезов на кромке излучателя Вивальди (рис. 4) на характеристики излучения рассмотрены вырезы высотой от 1 до 4 мм с шагом 1 мм и шириной от 1 до 4 мм с шагом 1 мм. Определено, что введение эллиптических вырезов на кромке шириной 1, 2 и 4 мм, а также высотой 1, 3 и 4 мм не приводит к значительному расширению рабочей полосы частот, однако снижает средний уровень КСВН. В итоге выбраны вырезы размером 3 × 2 мм, причем в нижней части излучателя ширина плавно возрастает до 10 мм, а их общее количество равно 35 с каждой стороны. Эллиптические вырезы на кромке расположены с периодом 4 мм.

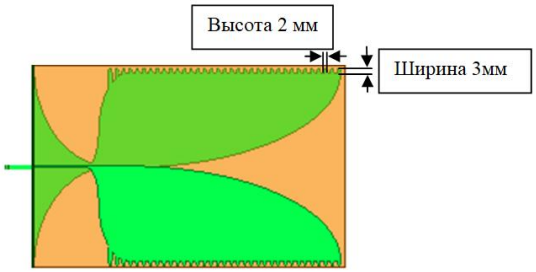
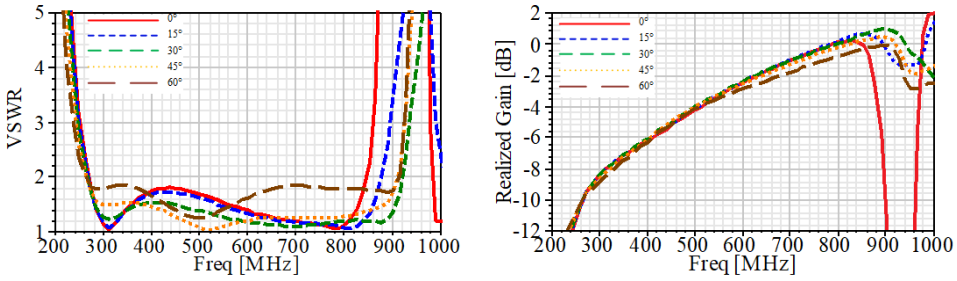



*Рис. 4. Излучатель Вивальди с эллиптическими вырезами на кромке*

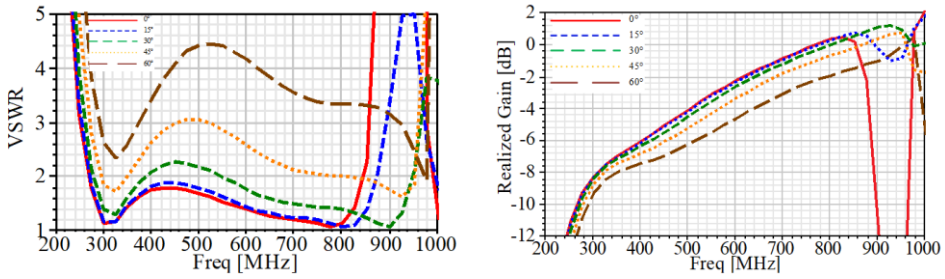
На рис. 5, 6 приведены характеристики излучателя Вивальди при размере эллиптических вырезов 3 × 2 мм (рис. 4), причем каждой кривой соответствует определенное значение угла сканирования: 0° (—), 15° (---), 30° (- -), 45° (·····) и 60° (— —).

Из рис. 5,а можно определить, что рабочая полоса частот излучателя Вивальди с эллиптическими вырезами на кромке в режиме сканирования в Е–плоскости в секторе углов ±60° находится в пределах от 251 до 857 МГц по уровню КСВН ≤ 3 (коэффициент перекрытия ~ 3,41). КУ в указанной полосе частот находится в пределах от минус 10,86 до 0,89 дБ, как видно из рис. 5,б.




а б

*Рис. 5. КСВН (а) и КУ (б) излучателя Вивальди с эллиптическими вырезами на кромке в режиме сканирования в Е–плоскости*

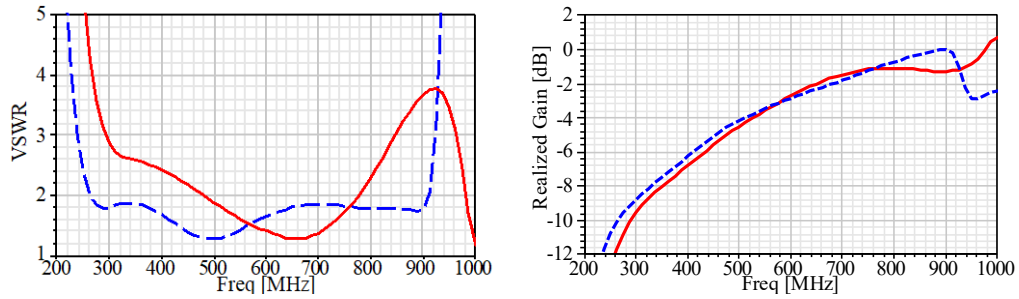



а б

*Рис. 6. КСВН (а) и КУ (б) излучателя Вивальди с эллиптическими вырезами на кромке в режиме сканирования в Н–плоскости*

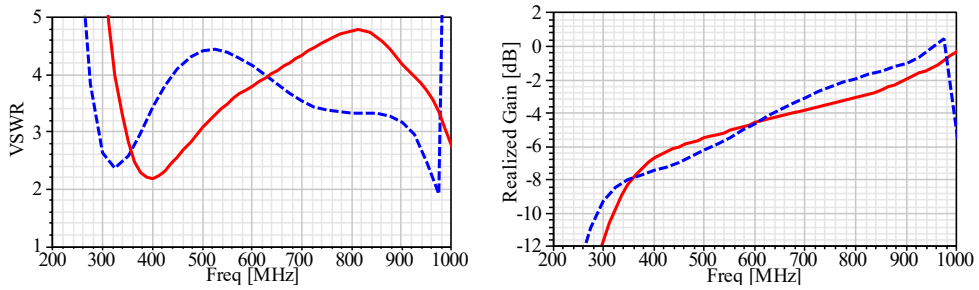
Согласно рис. 6,а, рабочая полоса частот излучателя Вивальди с эллиптическими вырезами на кромке в Н–плоскости находится в пределах от 271 до 854 МГц по уровню КСВН ≤ 3 (≤ 3,05 при 45°) в режиме сканирования в секторе углов ±45°. Коэффициент перекрытия ~ 3,15, а КУ в указанной полосе частот, как видно из рис. 6,б, находится в пределах от минус 11,27 до 0,59 дБ. При этом, уровень КСВН при сканировании на 45° на частотах 470–500 МГц располагается немного выше уровня КСВН = 3. Однако, если рассмотреть средний уровень КСВН в Е–плоскости – для излучателя с эллиптическими вырезами на кромке он окажется ниже на 0,8 по сравнению с излучателем без вырезов (рис. 2,а и 5,а).

Для наглядности влияния эллиптических вырезов на кромке излучателя на характеристики излучения при сканировании на 60° на рис. 7, 8 приведены КСВН и КУ для излучателя без вырезов (—) (рис. 1) и излучателя с эллиптическими вырезами на кромке (- - -) (рис. 4) в Е– и Н–плоскостях, соответственно.




а б

*Рис. 7. КСВН (а) и КУ (б) излучателя Вивальди без вырезов (—) и излучателя с эллиптическими вырезами на кромке (- - -) на 60° в Е–плоскости*




а б

*Рис. 8. КСВН (а) и КУ (б) излучателя Вивальди без вырезов (—) и излучателя с эллиптическими вырезами на кромке (---) на 60° в H–плоскости*

Из рис. 7,а можно определить, что рабочая полоса частот излучателя без вырезов на 60° в E–плоскости по уровню КСВН ≤ 2 находится в пределах от 480 до 776 МГц, а излучателя с эллиптическими вырезами на кромке по тому же уровню находится в пределах от 264 до 910 МГц. Рабочая полоса частот на 60° в E–плоскости при введении эллиптических вырезов на кромке излучателя расширяется на 350 МГц (с 296 МГц до 646 МГц). Увеличение КУ более чем на 1 дБ заметно в диапазоне от 850 до 920 МГц, согласно рис. 7,б.

Из рис. 8,а можно определить, что КСВН излучателя без вырезов и излучателя с эллиптическими вырезами на кромке на 60° в H–плоскости отличается по форме, однако его средний уровень не изменяется. КУ в диапазоне от 264 до 355 МГц и от 600 до 975 МГц увеличивается не более чем на 3,0 дБ, а в диапазоне от 355 до 600 МГц наоборот – уменьшается не более, чем на 1,2 дБ, согласно рис. 8,б.

Анализ проведенных исследований показывает, что рабочая полоса частот излучателя с эллиптическими вырезами на кромке по уровню КСВН ≤ 2 в E–плоскости оказывается шире и при сканировании до ±60° располагается от 270 до 842 МГц (коэффициент перекрытия ~ 3,12). В H–плоскости эллиптические вырезы на кромке позволяют расширить рабочую полосу частот на 28 МГц при сканировании в секторе углов ±45°.

Проведено сравнение разработанных излучателей без вырезов, с эллиптическими и прямоугольными вырезами на кромке [15]. Излучатель Вивальди с эллиптическими вырезами на кромке так же, как и излучатель без вырезов или излучатель с прямоугольной формой вырезов будет сверхширокополосным, однако средний уровень его КСВН лучше. Так, КСВН излучателя без вырезов в E–плоскости в основном располагается ниже уровня КСВН = 2,7 излучателя с прямоугольной формой вырезов – ниже уровня КСВН = 2,3, а излучателя с эллиптической формой вырезов – ниже уровня КСВН = 1,9. Более того, излучатель с эллиптическими вырезами на кромке будет являться СШП даже в режиме сканирования в секторе углов ±60° по уровню КСВН ≤ 2. КУ излучателей в рабочей полосе частот при введении вырезов разной формы изменяется слабо, так как размер ячейки бесконечной ФАР при использовании эллиптических вырезов на кромке не меняется.

Необходимо отметить, что электрические размеры разработанного СШП излучателя с эллиптическими вырезами на кромке в режиме сканирования в E–плоскости в секторе углов ±60° на верхней рабочей частоте 857 МГц составляют: 0,286 λ × 0,343 λ × 0,528 λ, а на нижней – 251 МГц: 0,084 λ × 0,100 λ × 0,155 λ.

**16-элементная антенная решетка.** Характеристики излучения, которые представлены на рис. 2, 3, 5 и 6, рассчитаны для излучателей в составе бесконечной ФАР. Для проверки работы излучателя без эллиптических вырезов и излучателя с эллиптическими вырезами в составе конечных решеток необходимо разработать электродинамическую модель АР конечного размера. Чем большее количество излучателей содержит АР, тем ближе будут характеристики излучения ее элементов к тем, которые представлены для излучателей в составе бесконечной ФАР. При этом, необходимо отметить, что большее количество излучателей АР конечного размера требует больших вычислительных затрат ЭВМ, вследствие чего были разработаны и рассчитаны электродинамические модели малоэлементных АР из 16 излучателей.

Исследованы АР с размером каждой ячейки 120 × 120 мм. Модель 16-элементной АР из излучателей без вырезов эллиптической формы на кромке из рис. 1, представлена на рис. 9. АР из излучателей с эллиптическими вырезами на кромке будет иметь такой же размер.

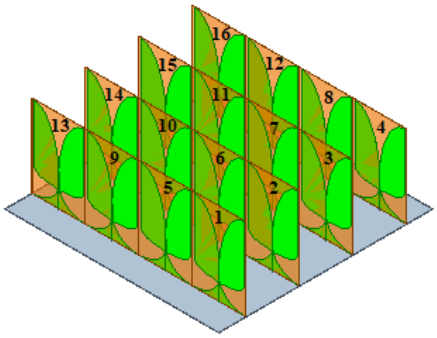


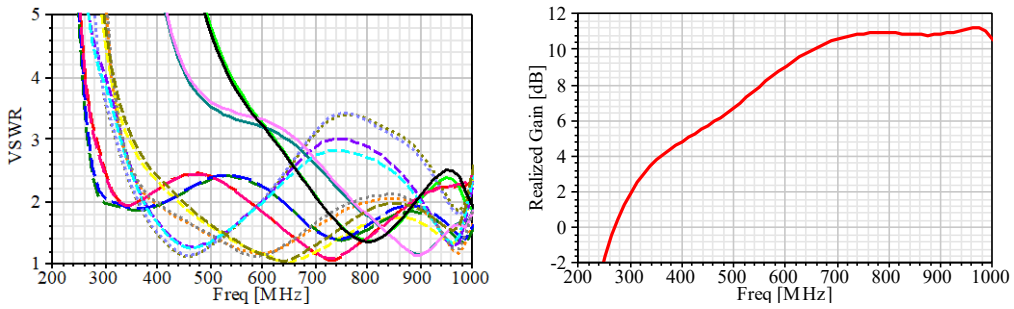

*Рис. 9. 16-элементная АР излучателей Вивальди без вырезов на кромке*

Характеристики излучения АР (см. рис. 9) при включении всех излучателей без вырезов представлены на рис. 10, а при подключении излучателей 1–4 на согласованные нагрузки сопротивлением 50 Ом представлены на рис. 11. В табл. 1 представлено соответствие между номером излучателя, а также цветом и формой кривой на графиках КСВН 16-элементной решетки. Данные таблицы справедливы как для АР излучателей без эллиптических вырезов, так и для АР с вырезами эллиптической формы на кромке.

Таблица 1

**Соответствие номера излучателя АР цвету и форме кривой**

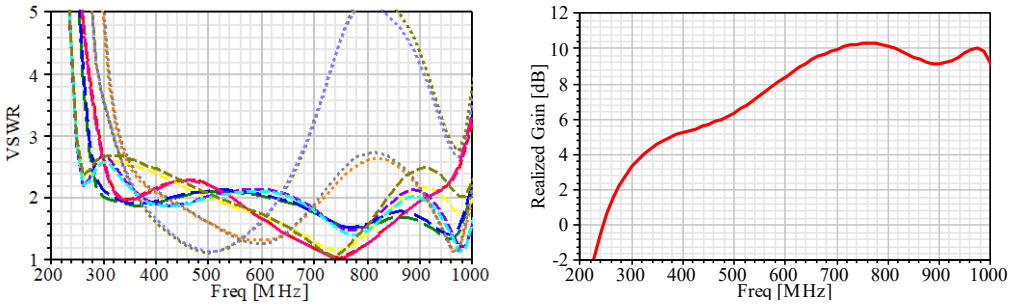
| Номер излучателя | Цвет и форма кривой | Номер излучателя | Цвет и форма кривой | Номер излучателя | Цвет и форма кривой | Номер излучателя | Цвет и форма кривой |
|---|---|---|---|---|---|---|---|
| 1 | | 5 | | 9 | | 13 | |
| 2 | | 6 | | 10 | | 14 | |
| 3 | | 7 | | 11 | | 15 | |
| 4 | | 8 | | 12 | | 16 | |




а б

*Рис. 10. КСВН (а) и КУ (б) 16-элементной решетки при включении всех излучателей без вырезов эллиптической формы на кромке*

При включении всех излучателей, представленных на рис. 9, ухудшается средний уровень КСВН всей решетки. Рабочая полоса частот по уровню КСВН ≤ 3 располагается от 872 до 1000 МГц, а коэффициент перекрытия ~ 1,15, как видно из рис. 10,а. Согласно рис. 10,б, КУ в указанной рабочей полосе частот находится в диапазоне от 10,76 до 11,19 дБ.

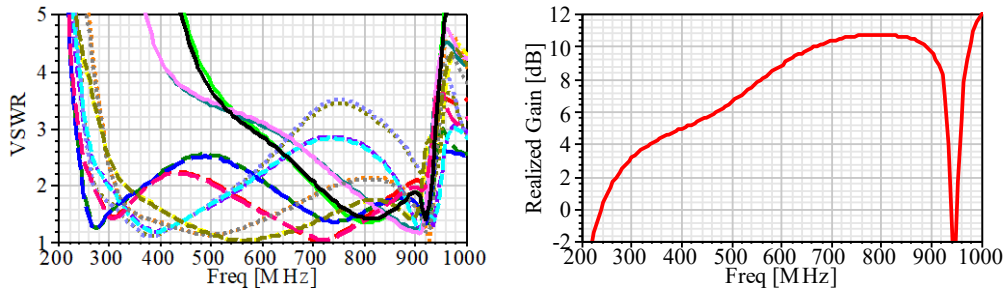



а б

*Рис. 11. КСВН (а) и КУ (б) 16-элементной решетки при подключении излучателей без вырезов 1–4 на согласованные нагрузки*

При этом, в случае подключения излучателей 1–4 (рис. 7) на согласованные нагрузки сопротивлением 50 Ом, средний уровень КСВН АР улучшается. Так, согласно рис. 11,а, рабочая полоса частот по уровню КСВН ≤ 3 располагается в диапазоне 332,6–688,2 МГц, чему соответствует коэффициент перекрытия ~ 2,07. КУ в рабочей полосе частот располагается в пределах от 4,28 до 9,83 дБ, что следует из рис. 11,б.

Характеристики излучения АР из излучателей с вырезами эллиптической формы на кромке (см. рис. 4) при включении всех излучателей представлены на рис. 12, а при подключении излучателей 1–4 на согласованные нагрузки сопротивлением 50 Ом представлены на рис. 13.

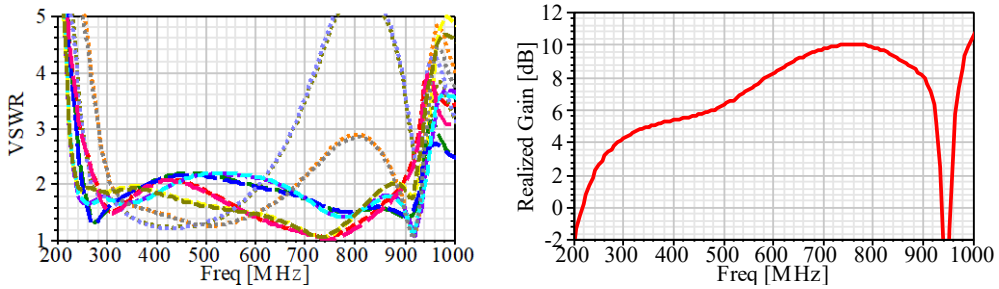



а б

*Рис. 12. КСВН (а) и КУ (б) 16-элементной решетки при включении всех излучателей с вырезами эллиптической формы на кромке*

При включении всех излучателей с эллиптическими вырезами ухудшается средний уровень КСВН всей решетки. Рабочая полоса частот по уровню КСВН ≤ 3 располагается от 485,8 до 943,6 МГц, а коэффициент перекрытия ~ 1,942, как видно из рис. 12,а. Согласно рис. 12,б, КУ в указанной рабочей полосе частот находится в диапазоне от минус 1,52 до 10,75 дБ.

В случае подключения излучателей с вырезами эллиптической формы на кромке 1–4 на согласованные нагрузки сопротивлением 50 Ом, средний уровень КСВН АР улучшается. Так, согласно рис. 13,а, рабочая полоса частот по уровню КСВН ≤ 3 в этом случае располагается в диапазоне 283,8–671,3 МГц, чему соответствует коэффициент перекрытия ~ 2,37. КУ в рабочей полосе частот располагается в пределах от 3,82 до 9,50 дБ, что следует из рис. 13,б.




а б

*Рис. 13. КСВН (а) и КУ (б) 16-элементной решетки при подключении излучателей с вырезами 1–4 на согласованные нагрузки*

Из рис. 11,б и 13,б видно, что при подключении излучателей 1–4 на согласованные нагрузки сопротивлением 50 Ом средний уровень КУ уменьшается. Это связано с уменьшением количества активных элементов АР. Таким образом, улучшение среднего уровня согласования АР связано с ухудшением среднего уровня КУ.

Таким образом, использование эллиптических вырезов на кромке излучателя Вивальди в составе 16-элементной АР позволяет расширить ее рабочую полосу частот на 330 МГц при включении всех излучателей и на 52 МГц при подключении излучателей 1–4 (см. рис. 9) на согласованные нагрузки сопротивлением 50 Ом. При этом, коэффициент перекрытия увеличивается с 2,07 до 2,37. Обе АР при подключении излучателей 1–4 на согласованные нагрузки сопротивлением 50 Ом будут СШП, однако коэффициент перекрытия АР излучателей с эллиптическими вырезами на кромке будет больше. КУ в случае подключения излучателей 1–4 АР на согласованные нагрузки сопротивлением 50 Ом практически не изменяется.

Как видно из представленных результатов, улучшение КСВН крайних элементов малоэлементных решеток является актуальной задачей и требует поиска новых путей решений этой проблемы. Одним из решений может являться подключение крайнего ряда излучателей на согласованные нагрузки, однако в этом случае наблюдается ухудшение КУ АР.

**Выводы.** Использование эллиптических вырезов на кромке размером 3 × 2 мм позволяется расширить рабочую полосу частот излучателя Вивальди в составе бесконечной ФАР в режиме сканирования в обеих плоскостях, а также улучшить согласование при работе в Е-плоскости. Так, в Е–плоскости в режиме сканирования в секторе углов ±60° коэффициент перекрытия по уровню КСВН ≤ 3 возрастает с 2,86 до 3,41, а в Н–плоскости в режиме сканирования в секторе углов ±45° коэффициент перекрытия по уровню КСВН ≤ 3,05 возрастает с 2,74 до 3,15. При этом, при работе в Е–плоскости в режиме сканирования в секторе углов ±60° излучатель является СШП как по уровню КСВН ≤ 3, так и по уровню КСВН ≤ 2, следовательно среднее значение согласования у излучателя с эллиптическими вырезами на кромке лучше.

Исследованы характеристики излучателя без эллиптических вырезов и с ними в составе разработанных 16-элементных АР. При добавлении эллиптических вырезов в конструкцию излучателей АР конечного размера коэффициент перекрытия увеличивается с 2,07 до 2,37. Обе АР при подключении излучателей 1–4 на согласованные нагрузки сопротивлением 50 Ом будут сверхширокополосными, однако коэффициент перекрытия АР излучателей с эллиптическими вырезами на кромке будет больше. Рабочий диапазон АР из излучателей с эллиптическими вырезами располагается в полосе частот 283,8–671,3 МГц, однако КСВН крайнего ряда излучателей 1–4 довольно велик. Необходимо искать новые пути решения этой проблемы, вследствие чего улучшение КСВН крайних элементов малоэлементных АР является актуальной задачей. Одним из решений может являться подключение крайнего ряда излучателей на согласованные нагрузки, однако в этом случае наблюдается ухудшение КУ АР.

*Работа выполнена при финансовой поддержке Российского научного фонда (Проект No22-19-00537, https://rscf.ru/project/22-19-00537/) в Центре коллективного пользования «Прикладная электродинамика и антенные измерения» Южного федерального университета, г. Таганрог.*

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Вендик О.Г., Парнес М.Д.* Антенны с электрическим сканированием (введение в теорию): учеб. пособие для вузов. – М.: Сайнс-пресс, 2002. – 231 с.
2. *Зырянов Ю.Т., Федюнин П.А., Белоусов О.А. [и др.]*. Антенны: учеб. пособие для вузов. – 5-е изд., стер. – СПб.: Лань, 2022. – 412 с.
3. *Гаврилов Л.Р.* Двумерные фазированные решетки для применения в хирургии: многофокусная генерация и сканирование // Акустический журнал. – 2003. – Т. 49, № 5. – С. 604-612.
4. *Гольдина И.М., Трофимова Е.Ю., Коков Л.С. [и др.]*. Возможности внутрисосудистого ультразвукового исследования с использованием катетерного датчика с фазированной решеткой в диагностике и лечении расслоения аорты // Ультразвуковая и функциональная диагностика. – 2016. – № 1. – С. 78-89.
5. *Калашников С.Н., Белоусов О.А.* Фазированная антенная решетка для систем цифрового телевидения // Вопросы современной науки и практики. Университет им. В.И. Вернадского. – 2014. – № 3 (53). – С. 62-67.
6. *Овчинникова Е.В., Гаджиев Э.В., Кондратьева С.Г. [и др.]*. Антенные решётки спутниковых мобильных телекоммуникационных систем // Вопросы электромеханики. Труды ВНИИЭМ. – 2021. – Т. 182, № 3. – С. 39-54.
7. *Хансен Р.С.* Сканирующие антенные системы СВЧ. Т. 1: пер. с англ. / под ред. Маркова Г.Т. и Чаплина А.Ф. – М.: Изд-во Советское радио, 1966. – 538 с.
8. *Воскресенский Д.И., Гостюхин В.Л., Максимов В.М. [и др.]*. Устройства СВЧ и антенны. Проектирование фазированных антенных решеток: учеб. пособие для вузов. – 3-е изд., доп. и перераб. – М.: Радиотехника, 2003. – 631 с.
9. *Gibson P.J.* The Vivaldi aerial // Proc. 9th European Microwave Conference. – 1979. – P. 101-105.
10. *Latha T., Ram G., Kumar G.A., Chakravarthy M.* Review on Ultra-Wideband Phased Array Antennas // IEEE Access. – 2021. – Vol. 9. – P. 129742-129755.
11. *Sahar Saleh, Mohd Haizal Jamaluddin, Faroq Razzaz, Saud M. Saeed, Nick Timmons, Jim Morrison* Compactness and performance enhancement techniques of ultra-wideband tapered slot antenna: A comprehensive review // Alexandria Engineering Journal. – 2023. – Vol. 74. – P. 195-229.
12. *Eichenberger J., Yetisir E., Ghalichechian N.* High-gain antipodal Vivaldi antenna with pseudoelement and notched tapered slot operating at (2.5 to 57) GHz // IEEE Trans. Antennas Propag. – 2019. – Vol. 67, No. 7. – P. 4357-4366.
13. *Dixit A.S., Kumar S.* A Survey of Performance Enhancement Techniques of Antipodal Vivaldi Antenna // IEEE Access. – 2020. – Vol. 8. – P. 45774-45796.
14. *Loo X.S., Win M.Z., Yeo K.S.* A high gain 60 GHz antipodal Fermi-tapered slot antenna based on robust synthesized dielectric // Microw. Opt. Technol. Lett. – 2018. – Vol. 61, No. 3. – P. 761-765.
15. *Kosak R.E., Gevorkyan A.V.* Research of Ways to Improve Radiation Characteristics of Phased Array Radiator Based on Vivaldi Antenna // 2021 Radiation and Scattering of Electromagnetic Waves (RSEMW). – 2021. – P. 211-214.
16. *Косак Р.Э., Геворкян А.В., Юханов Ю.В.* Излучатель фазированной антенной решетки узкоугольного сканирования // Компьютерные и информационные технологии в науке, инженерии и управлении (КомТех-2022). – 2022. – С. 258-263.
17. *Косак Р.Э., Геворкян А.В.* Компактный сверхширокополосный излучатель Вивальди кардиоидной формы с прямоугольными импедансными вставками // Известия ЮФУ. Технические науки. – 2024. – № 3. – С. 276-284. – DOI: 10.18522/2311-3103-2024-3-276-284.
18. *Briqech Z., Sebak A., Denidni T. A.* High gain 60 GHz antipodal Fermi tapered slot antenna with sine corrugation // Microw. Opt. Technol. Lett. – 2014. – Vol. 57, No. 1. – P. 6-9.
19. *Phalak K.D., Briqech Z., and Sebak A.* Ka-band antipodal Fermilinear tapered slot antenna with a knife edge corrugation // Microw. Opt. Technol. Lett. – 2014. – Vol. 57, No. 2. – P. 485-489.
20. *Инденбом М.В.* Антенные решетки подвижных обзорных РЛС. Теория, расчет, конструкции: монография. – М.: Радиотехника, 2015. – 416 с.

REFERENCES

1. *Vendik O.G., Parnes M.D.* Antenny s elektricheskim skanirovaniem (vvedenie v teoriyu): ucheb. posobie dlya vuzov [Electrically scanned antennas (introduction to theory): a textbook for universities]. Moscow: Sayns-press, 2002, 231 p.

2. *Zyryanov Yu.T., Fedyunin P.A., Belousov O.A. [i dr.].* Antenny: ucheb. posobie dlya vuzov [Antennas: a textbook for universities]. 5th ed. St. Petersburg: Lan', 2022, 412 p.
3. *Gavrilov L.R.* Dvumernye fazirovannye reshetki dlya primeneniya v khirurgii: mnogofokusnaya generatsiya i skanirovanie [2D phased arrays for surgical applications: multifocal generation and scanning], *Akusticheskiy zhurnal* [Acoustic magazine], 2003, Vol. 49, No. 5, pp. 604-612.
4. *Gol'dina I.M., Trofimova E.Yu., Kokov L.S. [i dr.].* Vozmozhnosti vnutrisosudistogo ul'tra-zvukovogo issledovaniya s ispol'zovaniem kateternogo datchika s fazirovannoy reshetkoy v diagnostike i lechenii rassloeniya aorty [The potential of intravascular ultrasound using a phased array catheter probe in the diagnosis and treatment of aortic dissection], *Ul'trazvukovaya i funktsional'naya diagnostika* [Ultrasound and functional diagnostics], 2016, No. 1, pp. 78-89.
5. *Kalashnikov S.N., Belousov O.A.* Fazirovannaya antennaya reshetka dlya sistem tsifrovogo televideniya [Phased array digital TV systems], *Voprosy sovremennoy nauki i praktiki. Universitet im. V.I. Vernadskogo* [Issues of modern science and practice. University named after V.I. Vernadsky], 2014, No. 3 (53), pp. 62-67.
6. *Ovchinnikova E.V., Gadzhiev E.V., Kondrat'eva S.G. [i dr.].* Antennye reshetki sputnikovykh mobil'nykh telekommunikatsionnykh sistem [Antenna arrays for satellite mobile telecommunication systems], *Voprosy elektromekhaniki. Trudy VNIIEM* [Questions of electromechanics. Proceedings of VNIIEM], 2021, Vol. 182, No. 3, pp. 39-54.
7. *Khansen R.S.* Skaniruyushchie antennye sistemy SVCh [Microwave scanning antennas]. Vol. 1: transl. from engl., ed. by Markov G.T. and Chaplin A.F. Moscow: Izd-vo Sovetskoe radio, 1966, 538 p.
8. *Voskresenskiy D.I., Gostyukhin V.L., Maksimov V.M. [i dr.].* Ustroystva SVCh i antenny. Proektirovanie fazirovannykh antennykh reshetok: ucheb. posobie dlya vuzov [Microwave devices and antennas. Design of phased array antennas: A textbook for universities]. 3rd ed. Moscow: Radiotekhnika, 2003, 631 p.
9. *Gibson P.J.* The Vivaldi aerial, *Proc. 9th European Microwave Conference,* 1979, pp. 101-105.
10. *Latha T., Ram G., Kumar G.A., Chakravarthy M.* Review on Ultra-Wideband Phased Array Antennas, *IEEE Access,* 2021, Vol. 9, pp. 129742-129755.
11. *Sahar Saleh, Mohd Haizal Jamaluddin, Faroq Razzaz, Saud M. Saeed, Nick Timmons, Jim Morrison* Compactness and performance enhancement techniques of ultra-wideband tapered slot antenna: A comprehensive review, *Alexandria Engineering Journal,* 2023, Vol. 74, pp. 195-229.
12. *Eichenberger J., Yetisir E., Ghalichechian N.* High-gain antipodal Vivaldi antenna with pseudoelement and notched tapered slot operating at (2.5 to 57) GHz, *IEEE Trans. Antennas Propag*., 2019, Vol. 67, No. 7, pp. 4357-4366.
13. *Dixit A.S., Kumar S.* A Survey of Performance Enhancement Techniques of Antipodal Vivaldi Antenna, *IEEE Access.*, 2020, Vol. 8, pp. 45774-45796.
14. *Loo X.S., Win M.Z., Yeo K.S.* A high gain 60 GHz antipodal Fermi-tapered slot antenna based on robust synthesized dielectric, *Microw. Opt. Technol. Lett*., 2018, Vol. 61, No. 3, pp. 761-765.
15. *Kosak R.E., Gevorkyan A.V.* Research of Ways to Improve Radiation Characteristics of Phased Array Radiator Based on Vivaldi Antenna, *2021 Radiation and Scattering of Electromagnetic Waves (RSEMW),* 2021, pp. 211-214.
16. *Kosak R.E., Gevorkyan A.V., Yukhanov Yu.V.* Izluchatel' fazirovannoy antennoy reshetki uzkougol'nogo skanirovaniya [Narrow-angle scanning phased array antenna radiator], *Komp'yuternye i informatsionnye tekhnologii v nauke, inzhenerii i upravlenii (KomTekh-2022)* [Computer and information technologies in science, engineering and management (KomTech-2022)], 2022, pp. 258-263.
17. *Kosak R.E., Gevorkyan A.V.* Kompaktnyy sverkhshirokopolosnyy izluchatel' Vival'di kardioidnoy formy s pryamougol'nymi impedansnymi vstavkami [Compact cardioid Vivaldi ultra-wideband radiator with rectangular impedance inserts], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2024, No. 3, pp. 276-284. DOI: 10.18522/2311-3103-2024-3-276-284.
18. *Briqech Z., Sebak A., Denidni T. A.* High gain 60 GHz antipodal Fermi tapered slot antenna with sine corrugation, *Microw. Opt. Technol. Lett*., 2014, Vol. 57, No. 1, pp. 6-9.
19. *Phalak K.D., Briqech Z., and Sebak A.* Ka-band antipodal Fermilinear tapered slot antenna with a knife edge corrugation, *Microw. Opt. Technol. Lett*., 2014, Vol. 57, No. 2, pp. 485-489.
20. *Indenbom M.V.* Antennye reshetki podvizhnykh obzornykh RLS. Teoriya, raschet, konstruktsii: monografiya [Antenna arrays of mobile surveillance radars. Theory, calculation, designs: monograph]. Moscow: Radiotekhnika, 2015, 416 p.

Статью рекомендовал к опубликованию д.т.н., профессор К.Е. Румянцев.

**Косак Роман Эдуардович** – Южный федеральный университет; email: kosak@sfedu.ru; г. Таганрог, Россия; тел.: +78634371733; кафедра АиРПУ; аспирант; зав. лабораторией.

**Kosak Roman Eduardovich** – Southern Federal University; email: kosak@sfedu.ru; Taganrog, Russia; phone: +78634371733; the Department of Antennas and Radiotransmitting Devices; postgraduate student, head of the laboratory.

УДК 004.272.2 DOI 10.18522/2311-3103-2024-5-243-254

**И.И. Левин, Д.С. Буряков**

# НЕКОТОРЫЕ МЕТОДЫ СИНХРОНИЗАЦИИ ИНФОРМАЦИОННЫХ ПОТОКОВ В СИСТЕМАХ ЦИФРОВОЙ ОБРАБОТКИ СИГНАЛОВ

*Предложены некоторые методы обеспечения когерентной обработки данных в системах радиолокации и связи, включающих фазированные антенные решетки (ФАР). Разработан подход для сбора оцифрованных данных от антенных элементов ФАР и передачи информации между распределенными узлами, которые выполняют цифровую обработку сигналов. Для обеспечения когерентной обработки и передачи данных предлагается использовать сигнал опорной тактовой частоты и единое машинное время, которые генерируются в центральном узле и распространяются по каналам с одинаковой задержкой. Все управляющие воздействия в узлах обработки основаны на данных сигналах. Для передачи оцифрованных данных от антенных элементов ФАР в работе предлагается использовать передачу фрагментами операндов с контролем целостности информации и привязкой ко времени оцифрованных данных. Проведенные эксперименты на реальном устройстве формирования диаграммы направленности подтвердили эффективность данного метода и его пригодность для практического использования. Развитие систем цифровой обработки сигналов (ЦОС) с ФАР постоянно движется вперед, требуется создание новых радиолокационных систем с высокой разрешающей способностью и достаточной чувствительностью. Обычно для повышения разрешающей способности увеличивают количество антенных элементов ФАР. Однако это приводит к увеличению размеров антенны и, следовательно, длины линий связи. При увеличении длины линии связи могут возникать различия в путях распространения сигнала из-за разброса характеристик оптических линий связи и воздействия внешних факторов на сигнал при его передаче через более длинные линии. Это может привести к неоднородности в задержках между каналами синхронизации и нарушению работы системы когерентной обработки. В связи с этим в работе предложен новый метод динамической компенсации задержек в каналах системы единого машинного времени для корректной работы с большими длинами линий связи.*

*Фазированная антенная решетка; программируемая логическая интегральная схема; когерентная обработка данных; система единого времени; компенсация задержек.*

**I.I. Levin, D.S. Buryakov**

# SOME METHODS FOR DATA FLOW SYNCHRONIZATION IN DIGITAL SIGNAL PROCESSING SYSTEMS

*The paper proposes some methods of providing coherent data processing in radar and communication systems that include phased arrays. An approach is developed to collect digitized data from the antenna elements of phased array and transfer information between distributed nodes that perform digital signal processing. To ensure coherent data processing and transmission, it is proposed to use a reference clock frequency signal and a common machine time, which are generated in the central node and distributed through the channels with the same delay. All control actions in the processing nodes are based on these signals. For the transmission of digitized data from the antenna elements of the phased array in the work it is proposed to use the transmission by fragments of operands with information integrity control and time reference of the digitized data. The experiments conducted on a real directional pattern formation device confirmed the effectiveness of this method and its suitability for practical use. The development of digital signal processing systems with phased array is constantly moving forward, and new radar systems with high resolution and sufficient sensitivity are required. Usually, to increase the resolution, the number of antenna elements is increased. However, this leads to an increase in the size of the antenna and hence the length of the links. As the length of the link increases, differences in signal propagation paths may occur due to the variation in the characteristics of optical links and the influence of external factors on the signal as it is transmitted through longer links. This can lead to inhomogeneity in delays between synchronization channels and disruption of the coherent processing system. In this regard, a new method of dynamic compensation of delays in the channels of the common machine time system for correct operation with long communication lines is proposed in this paper.*

*Phased array antenna; field programmable gate array; coherent data processing; common time system; delay compensation.*

**Введение.** Когерентная обработка данных применяется в задачах, предполагающих совместную, согласованную обработку информации, поступающей из различных источников со скоростью, сопоставимой с темпом поступления данных. На сегодняшний день подобная обработка сигналов получила широкое распространение и используется в различных областях электроники и радиотехники, включая такие важные области, как радиосвязь и радиолокация.

Одним из ключевых применений когерентной обработки является обработка данных, поступающих от фазированных антенных решеток. В сочетании с высокопроизводительными вычислительными устройствами ФАР позволяет формировать несколько независимо управляемых диаграмм направленности, обеспечивая точную локализацию объектов и эффективную работу радиосвязи. Кроме того, адаптивное формирование диаграмм направленности ФАР способствует снижению уровня помех, включая интерференцию от других источников, что особенно важно в условиях высоких электромагнитных помех, характерных для городской среды. Другим преимуществом ФАР является их способность работать при частичной деградации системы, например, при отказе отдельных антенных элементов [1].

Суть формирования диаграммы направленности в определенном направлении заключается в том, чтобы компенсировать временные задержки и суммировать сигналы от отдельных антенных элементов синфазно, то есть когерентно. Можно рассматривать диаграмму направленности как пространственный фильтр, который усиливает сигналы из заданного направления и подавляет сигналы из других направлений [2].

Основными параметрами диаграммы направленности фазированной антенной решетки являются ширина основного лепестка диаграммы направленности, уровень боковых лепестков, а также их пространственное направление относительно полотна антенной решетки. Данные параметры обеспечиваются за счёт необходимого фазового распределения по апертуре ФАР [3].

При цифровой обработке сигналов, поступающих от элементов ФАР, источником фазовых ошибок является разница фаз вследствие неодновременности оцифровки и обработки. Данная ошибка может возникать по двум основным причинам, первая из которых заключается в том, что отсчеты, полученные от различных аналогово-цифровых преобразователей (АЦП) данных от антенных элементов в один и тот же момент физического времени могут быть маркированы разными временными метками. Вторая причина связана с рассогласованием потоков данных во время их передачи к узлам обработки [4].

Для когерентной обработки данных критически важно обеспечить одновременное поступление данных, соответствующих одному моменту физического времени от всех без исключения антенных элементов ФАР. Далее по тексту такую передачу данных будем называть изохронной.

**Аппаратные средства когерентной обработки информационных потоков.** Обеспечение когерентной обработки данных может быть достигнуто с помощью специальных методов и средств обработки данных, а также более эффективных алгоритмов синхронизации работы всех узлов системы, которые будут рассмотрены в данной статье.

Вычислительные устройства высокой производительности для решения задач когерентной обработки информации от антенных элементов ФАР могут быть реализованы на различной элементной базе: на универсальных и специальных процессорах, специализированных интегральных схемах (ASIC) или программируемых логических интегральных схемах (ПЛИС).

Вычислители на основе универсальных процессоров являются гибкими и удобными средствами программирования. Однако, в контексте систем с обработкой информации от фазированных антенных решеток, они имеют несколько существенных недостатков, первым и наиболее значимым из которых является ограниченная вычислительная мощность универсальных процессоров. Алгоритмы обработки информации от ФАР часто требуют больших вычислительных затрат, т.к. включают в себя сложные и трудоемкие математические операции, расчеты и обработку больших объемов данных. В данном случае про-

изводительность процессоров может оказаться недостаточной для эффективной реализации подобных алгоритмов в режиме реального времени, что в дальнейшем может привести к высоким задержкам [5].

Вторым недостатком является недостаточная шина памяти. Для процессоров с общей шиной памяти доступ к памяти может быть узким местом при одновременном обращении нескольких потоков, что может снизить производительность системы.

Другим существенным недостатком является ограниченная параллельная обработка. В задачах обработки информации от фазированных антенных решеток требуется обработка больших объемов данных в режиме реального времени, что может быть трудно реализуемо на процессорах ввиду их ограниченной способности к параллельной обработке. Для эффективного использования ресурсов процессора требуется рациональное планирование и управление потоками. В случае многоканальной обработки данных могут возникать трудности в реализации в виду того, что процессору необходимо эффективно распределять ресурсы между несколькими потоками, что может вызывать задержки и увеличивать накладные расходы на организацию вычислительного процесса [6].

Наконец, еще одним недостатком универсальных процессоров является сложность их интеграции с аппаратными компонентами систем с ФАР. В системах с обработкой информации от ФАР часто требуется наличие развитой периферии для обеспечения множества быстрых каналов передачи когерентных данных между распределенными компонентами систем с ФАР, а также средств синхронизации для обеспечения когерентной обработки.

Альтернативой системам на основе универсальных процессоров в настоящее время являются системы, построенные на основе заказных интегральных схем специального назначения (ASIC). Главным недостатком использования заказных микросхем является их узкая направленность. Интегральные схемы специального назначения проектируются под конкретную прикладную задачу. Изготовление данных микросхем является трудоемким процессом: полный цикл разработки занимает от 12 до 18 месяцев. При изменении алгоритма решения задачи или в случае нахождения ошибок в схеме применять уже созданные микросхемы и специализированные устройства не представляется возможным, а для создания новых микросхем необходимо проводить все этапы разработки заново [7].

В связи с вышесказанным микросхемы ASIC целесообразно применять для создания изделий массового потребления. Специализированные вычислители, обрабатывающие информацию от ФАР, не являются массовым продуктом, поэтому применение заказных микросхем повлечет за собой существенное удорожание продукции.

В ряде работ [8, 9] описано применение реконфигурируемых вычислительных систем на основе ПЛИС в составе устройств, ориентированных на многоканальную высокопроизводительную обработку сигналов.

ПЛИС, подобно ASIC, проектируется под конкретную задачу, однако ПЛИС позволяют вносить корректировки алгоритмов работы, что обеспечивает гибкость устройств, построенных на их основе. Благодаря этой гибкости, можно быстро адаптировать систему к изменяющимся требованиям без необходимости переработки аппаратных средств. Наличие в ПЛИС большого числа ресурсов для построения различных алгоритмов цифровой обработки сигналов и множества внешних интерфейсов дает возможность организовать в системах ЦОС многопоточную когерентную обработку огромного количества данных. Также в ПЛИС возможно обеспечить минимальную задержку при обработке данных благодаря оптимизации логики и прямого доступа к аппаратным ресурсам. Это особенно критично для многоканальных систем в реальном времени, где даже небольшие задержки могут быть неприемлемы [10].

Применение ПЛИС для систем когерентной обработки информации от антенных элементов ФАР может сократить время и затраты на разработку за счет использования высокоуровневых языков программирования, более быстрой отладки и модификации функционала. Это делает разработку подобных систем на основе ПЛИС более эффективной и экономически целесообразной, в отличие от систем на основе универсальных процессоров или специализированных интегральных схем.

**Подсистема синхронизации информационных потоков.** Линии связи для передачи данных и сигналов синхронизации должны обладать низкими фазовыми шумами, высокой пропускной способностью, устойчивостью при работе в условиях воздействия электромагнитных помех [11]. В указанных условиях реализации линий связи для каналов синхронизации и передачи данных в комплексе ЦОС с ФАР наилучшим образом подходят оптические линии.

Как отмечалось ранее, одним из ключевых условий выполнения требований по обработке информации от ФАР является обеспечение изохронной передачи данных из различных распределенных источников для когерентной обработки в узлах формирования диаграмм направленности. Для этого необходимо обеспечить ряд согласованных событий по всем узлам системы, таких как одновременный запуск процесса оцифровки данных, временная привязка полученных цифровых данных, синхронизация вычислительных процессов формирования диаграмм направленности и другие.

Существует два подхода для обеспечения множества согласованных событий. Первый подход включает в себя создание всех необходимых сигналов для согласованной обработки в центральном узле и их распределение по всем узлам устройства формирования диаграмм направленности. Данный метод применим только в случае систем с небольшим количеством управляющих сигналов и узлов. Второй подход предполагает создание и распределение ограниченного числа опорных сигналов из центрального узла, а все необходимые сигналы для конкретного узла формируются внутри него (рис. 1). Для реализации второго подхода достаточно только двух сигналов: опорной тактовой частоты (ОТЧ) и сигнала единого машинного времени (EMB). Этот метод более предпочтителен, т.к. позволяет значительно сократить количество каналов передачи данных с одинаковой задержкой в системе [12].

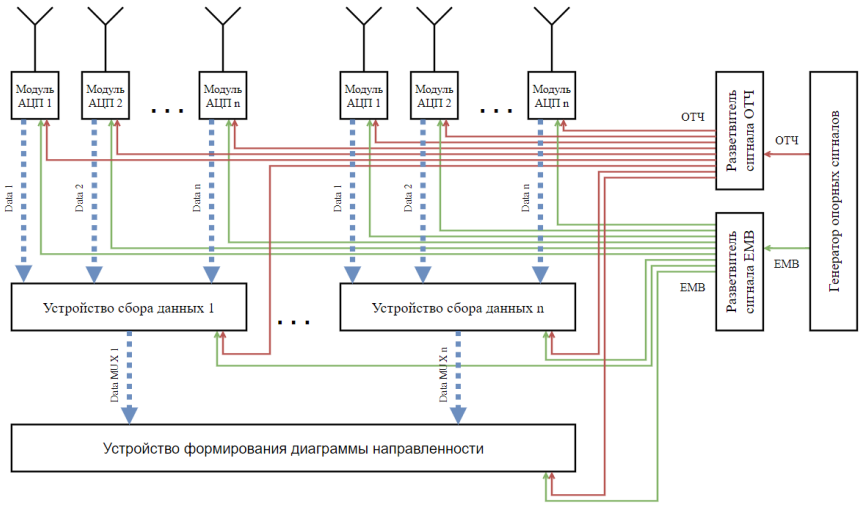



*Рис. 1. Структура системы ЦОС с обработкой информации от ФАР*

Единое машинное время и опорная тактовая частота распространяются по всем узлам из общего генератора опорных сигналов через линии связи одинаковой длины. Система машинного времени реализуется в каждом узле. У каждого узла есть счетчик текущего времени, который работает от опорной тактовой частоты и периодически обновляется значениями времени из источника единого машинного времени. Например, для запуска оцифровки данных требуется записать в модули АЦП время начала оцифровки. Когда время, указанное в счетчиках машинного времени, достигнуто, все модули АЦП одновременно начинают выполнение задачи. Аналогично запускаются все остальные процессы и генерируются управляющие сигналы, необходимые для когерентной обработки информации во всех узлах устройства формирования диаграмм направленности.

Использование единой опорной тактовой частоты для работы различных устройств обеспечивает синхронизацию их внутренних частотных схем. Это предотвращает «разбегание» внутренних частот устройств, вызванного различиями в характеристиках синтезаторов частот. Такой подход обеспечивает стабильность работы системы и предотвращает проблемы, связанные с различиями в работе внутренних синтезаторов частот [13].

Для передачи сигнала единого машинного времени на практике часто используется линейный самосинхронизирующийся код Манчестер-II [14]. Основная идея кода Манчестера заключается в кодировании каждого бита данных с использованием двух состояний сигнала, например, переход с низкого уровня на высокий или с высокого на низкий (рис. 2). Это позволяет обеспечить постоянное изменение состояния сигнала, что улучшает его устойчивость к шумам и помехам, а также обеспечивает синхронизацию. В коде Манчестера-II каждый бит данных кодируется с использованием двух переходов за период тактового сигнала. Таким образом, для передачи одного бита данных используется два тактовых сигнала. Это удвоенная версия оригинального кода Манчестера, которая обеспечивает дополнительную надежность передачи данных.

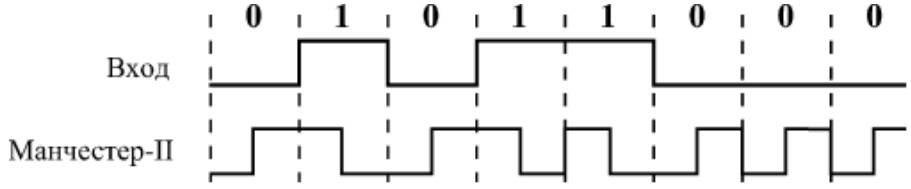



*Рис. 2. Временная диаграмма кода Манчестер 2*

Кроме того, Манчестерское кодирование не имеет постоянной составляющей, поскольку среднее значение сигнала равно нулю за счет изменения уровней сигнала на каждом такте. Отсутствие постоянной составляющей упрощает процесс декодирования сигнала на приемной стороне и делает сигнал менее чувствительным к постоянным помехам или смещениям уровня сигнала.

**Метод обеспечения согласованной передачи данных.** Для реализации гарантоспособности когерентной обработки необходимо обеспечить изохронную передачу оцифрованной информации по всем без исключения каналам данных в узлы обработки.

В тоже время, в каналах связи неизбежно возникают рассогласования, поскольку каналы передачи данных – это не только оптическая линия, но и преобразователи электрических сигналов в оптические и обратно, а также интерфейсные модули. Все функционирующие устройства вносят дополнительную задержку, которая в каждом из каналов может быть разной. Например, при проведении тренировки каналов интерфейсный модуль каждого канала, независимо от других, вставляет такты синхронизации. Рассогласование потоков данных между оптическими каналами является недопустимым, поскольку ведет к неправильному их использованию и, как следствие, некорректному результату обработки.

Прежде всего, требуется выровнять потоки данных в разных каналах для их согласованного получения в узлах приема. Кроме того, необходимо осуществить привязку оцифрованных данных к физическому времени.

Самый тривиальный способ передачи данных – применение непрерывного потока данных. Однако в этом случае проблематично выровнять потоки относительно друг друга и обеспечить привязку данных ко времени без существенных увеличений аппаратных затрат.

Классическая пакетная передача также не применима для решения данной задачи, поскольку требуется обеспечить поток данных в режиме реального времени, не допуская большой задержки. При классической пакетной передаче, после передачи каждого пакета, отправитель ожидает подтверждения от получателя. Если после определенного времени ожидания не было подтверждения получения пакета данных, то пакет отправляется снова [15]. Такой протокол создает недопустимые задержки передачи данных и может нарушить порядок следования пакетов данных, что недопустимо при когерентной обработке.

Поэтому был разработан гибридный способ передачи, сочетающий в себе элементы вышеперечисленных методов – передача данных потоком массивов операндов, разделенных служебными промежутками (рис. 3).

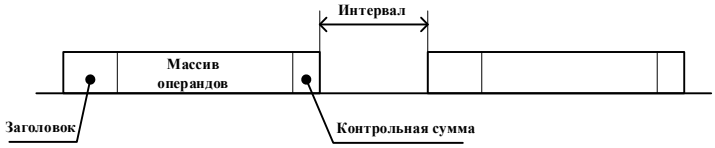



*Рис. 3. Передача данных массивами операндов*

Наличие служебных промежутков позволяет сформировать данные в неразрывные фрагменты, которыми можно легко манипулировать, включая выравнивание задержек между различными каналами передачи данных.

Одним из ключевых преимуществ разработанного подхода является возможность добавления к каждому фрагменту заголовка, содержащего сервисную информацию. Заголовок может содержать информацию о канале передачи, а также другие важные параметры, необходимые для корректной интерпретации и обработки данных на стороне приемника. Также в заголовке можно разместить временную метку, тем самым осуществив привязку ко времени всего массива операндов, но не каждого операнда в отдельности, минимизируя аппаратные и временные затраты. Более того, становится возможным внедрение средств контроля целостности передаваемой информации. Это позволяет обнаруживать и исправлять ошибки, возникающие в процессе передачи данных, что повышает надежность и точность передачи информации.

Для выравнивания задержек в различных каналах данных на стороне приемника требуется назначить один или несколько каналов в качестве опорных и выравнивать задержки в остальных каналах относительно опорных. Процесс выравнивания задержек всегда осуществляется относительно первого опорного канала, остальные каналы служат резервными на случай отказа текущего опорного канала. Все операции, корректирующие задержки, выполняются в интервале между массивами операндов.

Применение данного метода позволило обеспечить изохронную передачу информации в узлы обработки, а также повысить гарантоспособность подсистемы передачи данных.

Все представленные решения были апробированы в реальном устройстве формирования диаграммы направленности. Алгоритм выравнивания парировал возникающие задержки; сбоев в работе системы не отмечалось. Апробация на устройстве формирования диаграммы направленности подтвердила эффективность предложенных решений для практического применения. Согласно оценочным данным, гибридный метод может обеспечить гарантоспособность когерентной обработки для системы ЦОС с ФАР, включающую в себя 30000 антенных элементов с длинами межблочных связей более 100 метров.

**Ограничения применения метода.** Развитие систем с ФАР не стоит на месте. Возникают новые трудоемкие научно-технические задачи, к числу которых относится решение проблемы космического мусора, увеличение количества которого в последние годы стал серьезной угрозой безопасности космических полетов (рис. 4) [16]. С ростом коммерческих и государственных космических программ увеличивается количество запусков спутников и космических аппаратов. Каждый запуск добавляет новые объекты на орбиту, а отсутствие системы активного удаления мусора приводит к тому, что проблема становится все более острой. Все это создает риск столкновений, которые могут породить каскадный эффект, известный как «Синдром Кесслера» [17].

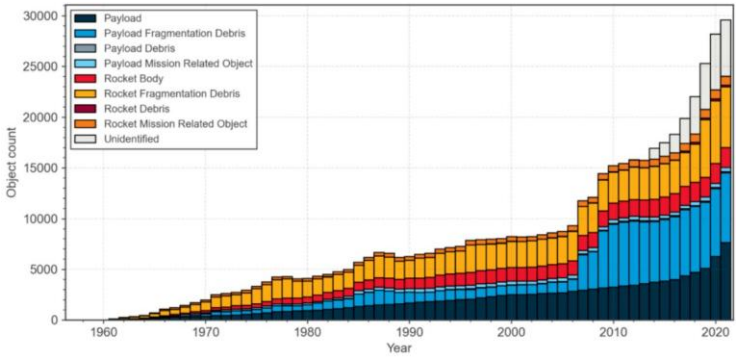



*Рис. 4. Динамика увеличения объемов космического мусора на околоземной орбите*

С увеличением объемов космического мусора возрастает актуальность и необходимость разработки эффективных средств обнаружения. Кроме того, на сегодняшний день отмечается необходимость обнаружения космических объектов меньшего размера, в отличие от тех, которые на данный момент могут быть обнаружены современными радиолокационными станциями. Как известно, даже малейшие фрагменты космического мусора, движущиеся со значительной скоростью, могут нанести значительный ущерб космическим аппаратам, находящимся на орбите. В связи с этим возникает необходимость в разработке новых радиолокационных систем с высокой разрешающей способностью и достаточной чувствительностью для обнаружения этих объектов.

Для повышения разрешающей способности, как правило, увеличивают количество антенных элементов ФАР, что позволяет сформировать более узкий луч за счет сужения диаграммы направленности [18]. Это позволит получить более точную информацию о положении и движении объектов в зоне обнаружения.

Следует отметить, что увеличение количества антенных элементов ФАР сопровождается рядом технических проблем, которые требуется решать на этапе разработки подобных систем. Прежде всего, увеличение количества антенных элементов приводит к увеличению апертуры антенны, а значит, и длин линий связи. Возникает проблема обеспечения когерентной обработки и изохронной передачи информации по каналам данных от мест приема в места обработки. При увеличении длины линии связи различия в путях распространения сигнала могут становиться более заметными из-за разброса характеристик различных оптических линий связи и влияния внешних факторов на сигнал при его передаче через более длинные линии [19]. Это может приводить к неоднородности в задержках между каналами, что неизбежно повлечет нарушение работы системы когерентной обработки.

Поэтому для обеспечения надлежащей работы системы с большими длинами линий связи необходимо применять новые подходы построения данных систем с ФАР.

**Метод динамической компенсации задержек в каналах системы цифровой обработки сигналов.** Основной проблемой при увеличении длины линий связи и возросшим при этом разбросе задержек в каналах данных является сложность в корректной доставке опорных сигналов. Это, в свою очередь, приводит к нарушению корректной работы системы машинного времени и появлению фазовых ошибок из-за обработки данных от АЦП, полученных в разные моменты физического времени.

Новый предлагаемый метод заключается в применении компенсации задержек, возникающих в каналах передачи сигналов системы EMB, с целью обеспечения синхронизации и корректной привязки оцифрованных данных к физическому времени.

Прежде всего, необходимо вычислить задержку от передатчика генератора EMB к приемнику EMB каждого узла. На данном этапе между узлами отсутствует синхронизация по времени, поэтому вычислить одностороннюю задержку (OWD – One-Way Delay) невозможно, для ее вычисления требуется одинаковое время в передатчике и приемнике. Поэтому вышеописанный способ передачи, использующий одностороннюю отправку сообщений EMB через разветвители к каждому узлу, не подходит.

Предлагается для системы EMB использовать двунаправленные каналы для передачи сигнала EMB из центрального узла к каждому устройству. Таким образом, становится возможным произвести вычисления круговой задержки (RTT – Round-Trip Time). Круговая задержка представляет собой время, затраченное на передачу данных от отправителя к получателю (OWD) и время, затраченное на передачу подтверждения получения данных обратно от получателя к отправителю [20].

Задачу вычисления и компенсации задержек в каналах EMB предлагается осуществлять в генераторе сигналов EMB, который является ключевым элементом в процессе обеспечения согласованности времени в системе. Такой подход обеспечивает централизованное управление процессом синхронизации и компенсации задержек, что упрощает управление и контроль, а также и повышает гарантоспособность системы синхронизации.

Предлагается реализовать два основных режима работы системы единого машинного времени для эффективной коррекции временных задержек. В первом режиме производится вычисление временных задержек для каждого канала передачи данных. Это позволяет установить точное значение задержки, которое затем используется для компенсации временных различий между различными каналами передачи. Во втором режиме работы системы происходит активное распространение времени с учетом ранее вычисленных временных задержек, что позволяет уравнять неоднородность в задержках линий связи. Структура системы EMB, использующая данный метод представлена на рис. 5.

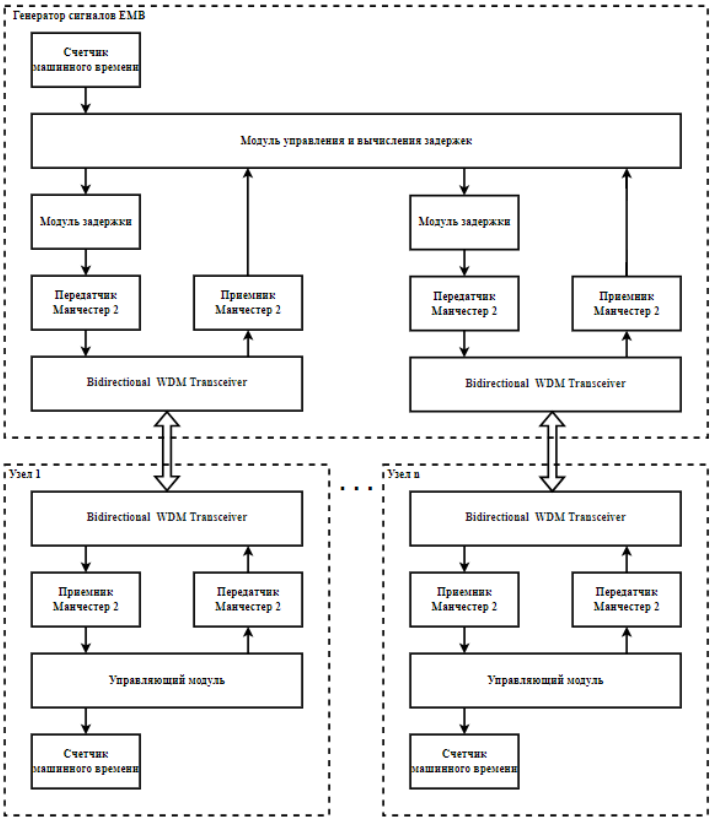



*Рис. 5. Структура системы единого машинного времени*

Рассмотрим подробнее процесс функционирования системы EMB (рис. 6).

В режиме вычисления временных задержек производится расчет круговой задержки (RTT). Для этого генератором сигналов EMB подается сигнал начала калибровки (start_calib) одновременно во все узлы, который заставляет приемники сигналов EMB узлов перейти в режим вычисления задержек. Далее генератор сигналов EMB отправляет

тестовый сигнал (test_signal), запуская при этом счетчики задержек для каждого канала в модуле управления и вычисления задержек. Каждый узел по получению тестового сигнала (test_signal) от генератора ЕМВ отправляет его обратно (test_echo). После получения сигнала test_echo модуль управления и вычислений задержек генератора ЕМВ останавливает соответствующие счетчики задержек по каждому каналу. Полученные значения передаются в модуль задержек, после чего генератор сигналов ЕМВ переключается в режим активного распространения времени. Модуль управления и вычисления задержек генератора сигналов ЕМВ отправляет сигнал сброса (reset), а затем передает значения машинного времени (machine_time) во все узлы с учетом вычисленных задержек.

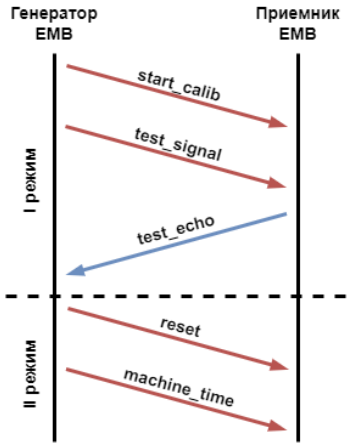



*Рис. 6. Работа системы единого машинного времени*

Отметим, что разработанный метод позволит нивелировать задержки между каналами в передаче сигнала ЕМВ, которые возникают из-за различия характеристик разных оптических линий с увеличением их протяженности. Таким образом будет обеспечена более точная синхронизация всех узлов системы, что позволит повысить гарантоспособность когерентной обработки. По предварительному анализу новый метод сможет обеспечить когерентную обработку информации от ФАР, состоящей из 100000 антенных элементов с длинами межблочных связей более 350 метров.

**Заключение.** В работе предложен и описан новый метод компенсации задержек, возникающих в каналах передачи сигналов системы единого машинного времени, с целью обеспечения синхронизации и корректной привязки оцифрованных данных к физическому времени.

Использование нового метода компенсации задержек позволит устранить неоднородности в задержках между каналами, вызванными разбросом параметров различных оптических линий при увеличении их длинны. Это, в свою очередь, позволит создавать более точные и чувствительные комплексы радиолокационная станций с ФАР, удовлетворяющие современным требованиям к радиолокационным системам в различных областях применения, в том числе для систем мониторинга околоземного космического пространства.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Воскресенский Д.И., Гостюхин В.О., Максимов В.М., Пономарев Л.И.* Устройства СВЧ и антенны / под ред. Д.И. Воскресенского. – 2-е изд., доп. и перераб. – М.: Радиотехника, 2006. – 376 с.
2. *Фомин А.Н., Тяпкин В.Н., Дмитриев Д.Д.* Теоретические и физические основы радиолокации и специального мониторинга: учебник / под ред. Ищук И.Н. – Красноярск: СФУ, 2016. – 292 с.
3. *Тепликова В.И., Сенцов А.А., Ненашев В.А., Поляков В.Б.* Анализ диаграммы направленности плоской многоэлементной активной фазированной антенной решетки // Тр. МАИ. – 2022. – № 125. – URL: https://cyberleninka.ru/article/n/analiz-diagrammy-napravlennosti-ploskoy-mnogoelementnoy-aktivnoy-fazirovannoy-antennoy-reshetki (дата обращения: 25.10.2024).

4. *Григорьев Л.Н.* Цифровое формирование диаграммы направленности в фазированных антенных решетках: монография. – М.: Радиотехника, 2010. – 144 с.
5. *Женетль С.Н., Куштанок С.А.* Современное развитие архитектур микропроцессоров // Новые технологии. – 2009. – № 2. – URL: https://cyberleninka.ru/article/n/sovremennoe-razvitie-arhitektur-mikroprotsessorov (дата обращения: 25.10.2024).
6. *Горобец А.В., Суков С.А., Триас Ф.Х.* Проблемы использования современных суперкомпьютеров при численном моделировании в гидродинамике и аэроакустике // Ученые записки ЦАГИ. – 2010. – № 2. – URL: https://cyberleninka.ru/article/n/problemy-ispolzovaniya-sovremennyh-superkompyuterov-pri-chislennom-modelirovanii-v-gidrodinamike-i-aeroakustike (дата обращения: 25.10.2024).
7. *Хаханов В.И., Обризан В.И., Мельникова О.В.* Обзор международного рынка электронных технологий // Вестник НТУ ХПИ. – 2004. – № 46. – URL: https://cyberleninka.ru/article/n/obzor-mezhdunarodnogo-rynka-elektronnyh-tehnologiy (дата обращения: 25.10.2024).
8. *Каляев И.А., Левин И.И., Семерников Е.А.* Высокопроизводительные реконфигурируемые вычислительные системы для цифровой обработки сигналов // Тр. Российского научно-технического общества радиотехники, электроники и связи имени А.С. Попова. Серия: Цифровая обработка сигналов и ее применение. – 2010. – Вып. XII – 1. – С. 13-18.
9. *Дордопуло А.И., Каляев И.А., Левин И.И., Семерников Е.А.* Высокопроизводительные многопроцессорные системы с реконфигурируемой архитектурой для цифровой обработки сигналов // Вестник Концерна ПВО «Алмаз-Антей». –2011. – № 2 (6). – С. 88-104.
10. *Чкан А.В.* Повышение реальной производительности РВС при решении задач цифровой обработки изображений с использованием быстрого преобразования Фурье // Известия ЮФУ. Технические науки. –2020. – № 7 (217). – URL: https://cyberleninka.ru/article/n/povyshenie-realnoy-proizvoditelnosti-rvs-pri-reshenii-zadach-tsifrovoy-obrabotki-izobrazheniy-s-ispolzovaniem-bystrogo (дата обращения: 25.10.2024).
11. *Куан И.А., Азимбаев Д.Ж., Щербаченя А.Н., Гербер А.С.* Волоконно-оптические линии связи // Вестник науки. –2018. – № 5 (5). – URL: https://cyberleninka.ru/article/n/volokonno-opticheskie-linii-svyazi (дата обращения: 25.10.2024).
12. *Савочкин И.А., Тройников Г.М., Тройникова Н.С., Турлаков П.В.* Система единого времени для высокоточной синхронизации разнесённых радиолокационных постов // Вестник Концерна ВКО «Алмаз – Антей». – 2014. – № 2. – С. 49-53.
13. *Сухман С.М., Бернов А.В., Шевкопляс Б.В.* Синхронизация в телекоммуникационных системах: Анализ инженерных решений. – М.: Эко-Трендз, 2002. – 268 с.
14. *Олифер В., Олифер Н.* Компьютерные сети. Принципы, технологии, протоколы: учебник для вузов. – 4-е изд. – СПб.: Питер, 2006. – 672 с.
15. *Пятибратов А.П., Гудыно Л.П., Кириченко А.А. и др.* П99 Вычислительные системы, сети и телекоммуникации: учебник. – 2-е изд., перераб. и доп. / под ред. А.П. Пятибратова. – М.: Финансы и статистика, 2004. – 512 с.
16. *Вениаминов С.С.* Космический мусор угрожает планете // Воздушно-космическая сфера. – 2016. – №1 (86). – URL: https://cyberleninka.ru/article/n/kosmicheskiy-musor-ugrozhaet-planete (дата обращения: 24.07.2024).
17. *Клюшников В.Ю.* Синдром Кесслера: будет ли закрыта дорога в космос? // ВКС. – 2021. – № 4 (109). – URL: https://cyberleninka.ru/article/n/sindrom-kesslera-budet-li-zakryta-doroga-v-kosmos (дата обращения: 25.10.2024).
18. *Дзюба А.П.* Перспективы развития фазированных антенных решеток // Вестник ДГТУ. Технические науки. – 2013. – № 3. – URL: https://cyberleninka.ru/article/n/perspektivy-razvitiya-fazirovannyh-antennyh-reshetok (дата обращения: 25.10.2024).
19. *Листвин А.В., Листвин В.Н.* Рефлектометрия оптических волокон. – М.: ЛЕСАРарт, 2005. - 208 с.
20. *Алешин В.С., Догаев С.Г.* Задержки распространения сигналов в сетях спутниковой связи // T-Comm. – 2019. – № 5. – URL: https://cyberleninka.ru/article/n/zaderzhki-rasprostraneniya-signalov-v-setyah-sputnikovoy-svyazi (дата обращения: 27.10.2024).

REFERENCES

1. *Voskresenskiy D.I., Gostyukhin V.O., Maksimov V.M., Ponomarev L.I.* Ustroystva SVCh i antenny [UHF devices and antennas], ed. by D.I. Voskresenskogo. 2nd ed. Moscow: Radiotekhnika, 2006, 376 p.
2. *Fomin A.N., Tyapkin V.N., Dmitriev D.D.* Teoreticheskie i fizicheskie osnovy radiolokatsii i spetsial'nogo monitoringa: uchebnik [Theoretical and physical basis of radar and special monitoring: textbook], ed. by Ishchuk I.N. Krasnoyar: SFU, 2016, 292 p.

3. *Teplikova V.I., Sentsov A.A., Nenashev V.A., Polyakov V.B.* Analiz diagrammy napravlennosti ploskoy mnogoelementnoy aktivnoy fazirovannoy antennoy reshetki [Directional pattern analysis of a planar multi-element active phased array antenna array], *Tr. MAI* [Trudy MAI], 2022, No. 125. Available at: https://cyberleninka.ru/article/n/analiz-diagrammy-napravlennosti-ploskoy-mnogoelementnoy-aktivnoy-fazirovannoy-antennoy-reshetki (accessed 25 October 2024).
4. *Grigor'ev L.N.* TSifrovoe formirovanie diagrammy napravlennosti v fazirovannykh antennykh reshetkakh: monografiya [Digital directional pattern formation in phased antenna arrays: monograph]. Moscow: Radiotekhnika, 2010, 144 p.
5. *Zhenetl' S.N., Kushtanok S.A.* Sovremennoe razvitie arkhitektur mikroprotsessorov [Modern development of microprocessor architectures], *Novye tekhnologii* [New technologies], 2009, No. 2. Available at: https://cyberleninka.ru/article/n/sovremennoe-razvitie-arhitektur-mikroprotsessorov (accessed 25 October 2024).
6. *Gorobets A.V., Sukov S.A., Trias F.Kh.* Problemy ispol'zovaniya sovremennykh superkomp'yuterov pri chislennom modelirovanii v gidrodinamike i aeroakustike [Problems of using modern supercomputers in numerical modeling in hydrodynamics and aeroacoustics], *Uchenye zapiski TsAGI* [Scientific notes of TsAGI], 2010, No. 2. Available at: https://cyberleninka.ru/article/n/problemy-ispolzovaniya-sovremennyh-superkompyuterov-pri-chislennom-modelirovanii-v-gidrodinamike-i-aeroakustike (accessed 25 October 2024).
7. *Khakhanov V.I., Obrizan V.I., Mel'nikova O.V.* Obzor mezhdunarodnogo rynka elektronnykh tekhnologiy [An overview of the international market for electronic technologies], *Vestnik NTU KhPI* [Bulletin of the National Technical University Kharkov Polytechnic Institute], 2004, No. 46. Available at: https://cyberleninka.ru/article/n/obzor-mezhdunarodnogo-rynka-elektronnyh-tehnologiy (accessed 25 October 2024).
8. *Kalyaev I.A., Levin I.I., Semernikov E.A.* Vysokoproizvoditel'nye rekonfiguriruemye vychislitel'nye sistemy dlya tsifrovoy obrabotki signalov [High-performance reconfigurable computing systems for digital signal processing], *Tr. Rossiyskogo nauchno-tekhnicheskogo obshchestva radiotekhniki, elektroniki i svyazi imeni A.S. Popova. Seriya: Tsifrovaya obrabotka signalov i ee primenenie* [Proceedings of the Russian Scientific and Technical Society of Radio Engineering, Electronics and Communications named after A.S. Popov. Series: Digital signal processing and its application], 2010, Issue XII – 1, pp. 13-18.
9. *Dordopulo A.I., Kalyaev I.A., Levin I.I., Semernikov E.A.* Vysokoproizvoditel'nye mnogoprotsessornye sistemy s rekonfiguriruemoy arkhitekturoy dlya tsifrovoy obrabotki signalov [High-performance multiprocessor systems with reconfigurable architecture for digital signal processing], *Vestnik Kontserna PVO «Almaz-Antey»* [Bulletin of the Almaz-Antey Air Defense Concern], 2011, No. 2 (6), pp. 88-104.
10. *Chkan A.V.* Povyshenie real'noy proizvoditel'nosti RVS pri reshenii zadach tsifrovoy obra-botki izobrazheniy s ispol'zovaniem bystrogo preobrazovaniya Fur'e [Improving the real performance of RCS in solving digital image processing problems using the fast Fourier transform], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2020, No. 7 (217). Available at: https://cyberleninka.ru/article/n/povyshenie-realnoy-proizvoditelnosti-rvs-pri-reshenii-zadach-tsifrovoy-obrabotki-izobrazheniy-s-ispolzovaniem-bystrogo (accessed 25 October 2024).
11. *Kuan I.A., Azimbaev D.Zh., Shcherbachenya A.N., Gerber A.S.* Volokonno-opticheskie linii svyazi [Fiber-optic communication lines], *Vestnik nauki* [Bulletin of science], 2018, No. 5 (5). Available at: https://cyberleninka.ru/article/n/volokonno-opticheskie-linii-svyazi (accessed 25 October 2024).
12. *Savochkin I.A., Troynikov G.M., Troynikova N.S., Turlakov P.V.* Sistema edinogo vremeni dlya vysokotochnoy sinkhronizatsii raznesennykh radiolokatsionnykh postov [A unified time system for high-precision synchronization of distributed radar posts], *Vestnik Kontserna VKO «Almaz – Antey»* [Bulletin of the Almaz-Antey Air Defense Concern], 2014, No. 2, pp. 49-53.
13. *Sukhman S.M., Bernov A.V., Shevkoplyas B.V.* Sinkhronizatsiya v telekommunikatsionnykh sistemakh: Analiz inzhenernykh resheniy [Synchronization in telecommunication systems: Analysis of engineering solutions]. Moscow: Eko-Trendz, 2002m 268 p.
14. *Olifer V., Olifer N.* Komp'yuternye seti. Printsipy, tekhnologii, protokoly: uchebnik dlya vuzov [Computer networks. Principles, technologies, protocols: textbook for universities]. 4th ed. Saint Petersburg: Piter, 2006, 672 p.
15. *Pyatibratov A.P., Gudyno L.P., Kirichenko A.A. i dr.* P99 Vychislitel'nye sistemy, seti i telekommunikatsii: uchebnik [Computer systems, networks and telecommunications: textbook]. 2nd ed., ed. by A.P. Pyatibratova. Moscow: Finansy i statistika, 2004, 512 p.
16. Veniaminov S.S. Kosmicheskiy musor ugrozhaet planete [Space debris threatens the planet], *Vozdushno-kosmicheskaya sfera* [Air and space sphere], 2016, No. 1 (86). Available at: https://cyberleninka.ru/article/n/kosmicheskiy-musor-ugrozhaet-planete (accessed 24 July 2024).

17. *Klyushnikov V.Yu.* Sindrom Kesslera: budet li zakryta doroga v kosmos? [Kessler syndrome: will the road to space be closed?], *VKS* [Aerospace sphere], 2021, No. 4 (109). Available at: https://cyberleninka.ru/article/n/sindrom-kesslera-budet-li-zakryta-doroga-v-kosmos (accessed 25 October 2024).
18. *Dzyuba A.P.* Perspektivy razvitiya fazirovannykh antennykh reshetok [Prospects for the development of phased antenna arrays], *Vestnik DGTU. Tekhnicheskie nauki* [Bulletin of DSTU. Technical sciences], 2013, No. 3. Available at: https://cyberleninka.ru/article/n/perspektivy-razvitiya-fazirovannyh-antennyh-reshetok (accessed 25 October 2024).
19. *Listvin A.V, Listvin V.N.* Reflektometriya opticheskikh volokon [Reflectometry of optical fibers]. Moscow: LESARart, 2005, 208 p.
20. *Aleshin V.S., Dogaev S.G.* Zaderzhki rasprostraneniya signalov v setyakh sputnikovoy svyazi [Signal propagation delays in satellite communication networks], *T-Comm,* 2019, No. 5. Available at: https://cyberleninka.ru/article/n/zaderzhki-rasprostraneniya-signalov-v-setyah-sputnikovoy-svyazi (accessed 27 October 2024).

Статью рекомендовал к опубликованию д.т.н., профессор И.И. Турулин.

**Левин Илья Израилевич** – Южный федеральный университет; e-mail: levin@superevm.ru; г. Таганрог, Россия; тел.: +78634612111; кафедра интеллектуальных и многопроцессорных систем, зав. кафедрой интеллектуальных и многопроцессорных систем; д.т.н.; профессор.

**Буряков Дмитрий Сергеевич** – e-mail: dburiakov@sfedu.ru; тел.: +79198955502; аспирант.

**Levin Ilya Izrailevich** – Southern Federal University; e-mail: levin@superevm.ru; Taganrog, Russia; phone: +78634612111; the Department of Intelligent and Multiprocessor Systems; head of Department; dr. of eng. sc.; professor.

**Buryakov Dmitrii Sergeevich** – e-mail: dburiakov@sfedu.ru; phone: +79198955502; postgraduate student.

УДК 621.396.624 DOI 10.18522/2311-3103-2024-5-254-260

**А.П. Плёнкин**

## СПОСОБ ОБНАРУЖЕНИЯ ОПТИЧЕСКОГО СИГНАЛА В КВАНТОВЫХ СЕТЯХ

*Приводится способ обнаружения оптического сигнала синхронизации для участка сети квантовых коммуникаций. Целью статьи является представление варианта реализации городской квантовой сети. В работе рассматривается решение задачи конфигурации канала синхронизации для систем квантовой связи нестандартной топологии. Описывается обобщенный принцип работы системы квантового распределения ключей с фазовым кодированием. Предлагается алгоритм синхронизации, адаптированный для конфигурации городской квантовой сети, содержащей несколько сегментов. Особенностью предлагаемой схемы является наличие одной приемо-передающей станции, с которой взаимодействуют несколько кодирующих станций. В статье приведены результаты анализа энергетической модели предлагаемого способа и расчет усредненных потерь в квантовом канале. В заключении мы рассуждаем о возможных вариантах структуры квантовых сетей и о применимости в них процессов синхронизации. Сети квантовых коммуникаций активно масштабируются и используют различные протоколы квантового распределения ключей, аутентификации и синхронизации. Квантовое распределение ключей (КРК) решает центральную проблему симметричной криптографии и представляет собой безопасную технологию генерации идентичной последовательности бит у двух удаленных пользователей. Теоретически, безопасность (стойкость) такой технологии не зависит от вычислительной мощности взломщиков, которые, например, могут обладать квантовым компьютером. Тем не менее, практическая реализация теоретических моделей все еще показывает техническое несовершенство, что позволяет злоумышленникам находить уязвимости. При исследовании и проектировании различных модификаций систем квантового распределения ключей (СКРК), необходимо уделять внимание не только вопросам стойкости квантовых протоколов, но и компонентам технической реализации аппаратуры.*

*Квантовые коммуникации; квантовый ключ; фотонный импульс; вероятность обнаружения; доверенные узлы.*

**A.P. Pljonkin**

## METHOD FOR DETECTING OPTICAL SIGNAL IN QUANTUM NETWORKS

*The article presents a method for detecting an optical synchronization signal for a section of a quantum communications network. The objective of the article is to present a variant of implementing an urban quantum network. The paper considers a proposed solution to the problem of configuring a synchronization channel for quantum communication systems with a non-standard topology. A generalized operating principle of a quantum key distribution system with phase coding is described. A synchronization algorithm adapted for configuring an urban quantum network containing several segments is proposed. A feature of the proposed scheme is the presence of one receiving and transmitting station with which several coding stations interact. The article presents the results of analyzing the energy model of the proposed method and calculating the average losses in the quantum channel. In conclusion, we discuss possible variants of the structure of quantum networks and the applicability of synchronization processes in them. Quantum communications networks are actively scaling and use various quantum key distribution, authentication, and synchronization protocols. Quantum key distribution (QKD) solves the central problem of symmetric cryptography and is a secure technology for generating an identical bit sequence for two remote users. Theoretically, the security (resistance) of such technology does not depend on the computing power of hackers, who, for example, may have a quantum computer. However, the practical implementation of theoretical models still shows technical imperfection, which allows attackers to find vulnerabilities. When researching and designing various modifications of quantum key distribution systems (QKDS), it is necessary to pay attention not only to the issues of the stability of quantum protocols, but also to the components of the technical implementation of the equipment.*

*Quantum communications; quantum key; photon pulse; detection probability; trusted nodes.*

**Введение.** В сетях квантовых коммуникаций базовой конфигурацией является топология «точка-точка», при которой квантовые секретные ключи распределяются между двумя удаленными узлами. Ограничения такой топологии заключаются, в том числе, возможным предельным расстоянием. Максимальное эффективное расстояние связано с особенностями функционирования квантовых протоколов. В подавляющем большинстве для работы протоколов квантового распределения ключей требуются ослабленные до однофотонного уровня оптические сигналы. Протяженные волоконно-оптические линии связи вносят существенное затухание и не позволяют передавать слабый оптический сигнал на большие расстояния без усиления. Смешанные топологии квантово-криптографических сетей сегодня построены на масштабировании базовой конфигурации и требуют наличия общего ключа между любой парой узлов, в том числе тех, которые не связаны непосредственно квантовым каналом связи. Задача распределения секретных ключей решается применением доверенных промежуточных узлов (ДПУ), через которые по цепочке передаются ключи к необходимым нодам сети. В Китае по такому принципу построена сложная квантовая сеть, охватывающая десятки городов и имеющая протяженность в тысячи километров [1]. В России также используют подход с ДПУ при построении сетей квантовых коммуникаций. Конструктивно ДПУ представляет собой безопасное помещение с оборудованием квантовой криптографией. В большинстве действующих протоколах квантового распределения ключей секретная последовательность формируется путем срабатывания нескольких лавинных фотодиодов (ЛФД) [2–4]. Например, срабатывание одного ЛФД интерпретируется как «0», а срабатывание другого как «1». Технической задачей при проектировании ДПУ является запрет доступа злоумышленника к оборудованию КРК, так как доступ к ЛФД позволит получить необходимую информацию о квантовой последовательности.

Известны модификации систем КРК, которые позволяют вынести ЛФД в отдельное неконтролируемое пространство. Исследования показывают, что в этом случае злоумышленник может иметь полный доступ к детекторам и это не влияет на секретность квантового протокола. Подобная модификация может быть применима лишь в отдельно взятых случаях, так как доступ к системе КРК все равно должен быть ограничен. В последнее десятилетие активно исследуются методы КРК на перепутанных парах фотонов (TF QKD) и с недоверенными промежуточными узлами (НПУ). В таких недоверенных

узлах допускается, что злоумышленник обладает всей информацией о работе аппаратуры, включая работу ЛФД. Система квантовой связи с НДУ базируется на топологии «точка-точка» с центральной недоверенной нодой. Квантовое распределение в такой сети реализуется по протоколу MDI (Measurement Device Independent).

В работе [5] описывается доказательство стойкости протокола MDI, принцип которого схож с известным BB84. Отправитель и получатель равновероятно выбирают один из базисов. Эта процедура происходит независимо друг от друга. Далее аналогично происходит выбор ортогонального состояния и присваиваются значения 0 и 1. Состояния поступают на НДУ, где производятся измерения в неполном Белловском базисе [6, 7]. Отметим, что результаты измерений на недоверенном узле общедоступны. Далее посылки, в которых использованы разные базисы, отбрасываются, а респонденты производят операции инвертации бит. В такой схеме требуется оценка вероятности ошибки, которая позволяет определить величину утечки информации к нарушителю.

**Подготовительные процессы КРК.** Работа квантового протокола является одной из финальных стадий в операциях систем КРК, функционирование которых невозможно без предварительных процедур настройки и согласования. Квантовые сети в базовой топологии «точка-точка» содержат три канала связи между отправителем и получателем: квантовый, синхронизации, общедоступный. Квантовый канал – это оптический тракт (оптическое волокно или оптический атмосферный канал), по которому реализуется работа квантового протокола. Канал синхронизации (или калибровки) – в большинстве случаев это отдельный волоконно-оптический канал для согласования и периодической подстройки компонентов системы КРК. Квантовый канал и синхронизация могут быть совмещены, т.е. физически реализованы в одном оптическом волокне [8]. Общедоступный канал – это сеть передачи данных, по которой осуществляются процессы аутентификации, шифрования, дешифрования.

Сегодня существует множество реализаций систем КРК, но принцип действия и ряд компонентов у всех схожий. Как правило, всегда есть ЛФД, источник оптического излучения, интерферометры, фазовые модуляторы, поляризационные фильтры. Рассмотрим двухпроходную схему реализации СКРК, в которой оптические ЛФД, источник излучения и интерферометр Маха-Цендера расположены в одной станции (Алиса) [9]. В такой системе станции Алиса и Боб соединены одним оптическим волокном, по которому реализуется синхронизация и работа квантового протокола. Удобство такой схемы заключается в том, что все технологически сложные элементы расположены в одном модуле (корпусе). При построении квантовых сетей такая конфигурация может быть востребована, когда требуется распределять ключи между базовой станцией и пользователями (рис. 1). В подобных схемах наиболее эффективным является использование квантового протокола BB84 и его модификаций с фазовым или поляризационным кодированием.

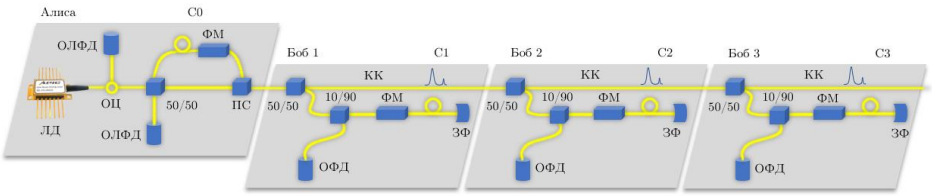



*Рис. 1. Схема сети квантовых коммуникаций*

Предположим, что в нашем распоряжении есть «темное» оптическое волокно, расположенное вдоль всех сегментов сети. Такое волокно служит как для передачи квантовых сигналов (КК), так и для синхронизации. Перед началом работы и в процессе функционирования квантового протокола системе КРК необходима калибровка. Так как схема двухпроходная, то оптическим сигналам необходимо пройти путь от источника излучения (ЛД) до вращающего поляризацию зеркала Фарадея (ЗФ) и обратно в станцию Алиса к ОЛФД. Отметим, что практическая реализация двухпроходной схемы имеет сущест-

венные ограничения по расстоянию, но вполне применима для городских задач с дистанцией в 50-70 км. Рассмотрим процесс калибровки, задача которого заключается в определении точного расстояния от ЛД до фазового модулятора и ОЛФД. Системе КРК необходимо понимать, в какой момент времени прикладывать электрическое напряжение к ФМ (Боб 1) и в какой момент времени подавать сигнал на ОЛФД. Наиболее распространенный процесс измерения расстояния через обнаружение оптического синхросигнала описан в [10]. В двухпроходных системах КРК данный метод применяется с небольшими модификациями.

Периодическая последовательность оптических импульсов (*1550 нм*) следует на делитель мощности (50/50) через циркулятор (ОЦ) и распределяется по плечам интерферометра. Каждый импульс последовательности разделяется на два с временной задержкой, равной длине отрезка волокна перед фазовым модулятором (ФМ) в плече интерферометра. Далее поляризационный сплиттер направляет сигналы в квантовый канал. По достижению первого сегмента квантовой сети, часть каждого импульса отводится через делитель мощности в станцию Боб 1. Здесь, пройдя ряд волоконно-оптических элементов, сигналы фиксируются классическим детектором (ОФД) и отражаются от ЗФ, изменяя поляризацию на ортогональную. При обратном распространении синхросигналы интерферируют на светоделителе и фиксируются ОЛФД.

Часть сигналов минует сегмент «Боб 1» и направляется по каналу к следующему сегменту «Боб 2», где происходит подобная (как в Боб 1) операция с последовательностью импульсов. Аналогичный процесс применим к последующим сегментам. Так как в процессе синхронизации ОЛФД работают в линейном режиме, то им не требуется время для восстановления работоспособности после детектирования. При расчете числа сегментов сети следует учитывать максимальный период следования импульсов таким образом, чтобы отраженный в последнем модуле оптический синхроимпульс мог вернуться в станцию отправителя и не пересечься по пути с встречным импульсом. В реализованных система КРК период следования тактовых импульсов равен *1.2 мс*. Такой период рассчитан, исходя из максимально возможного расстояния между удаленными станциями (рис. 2).

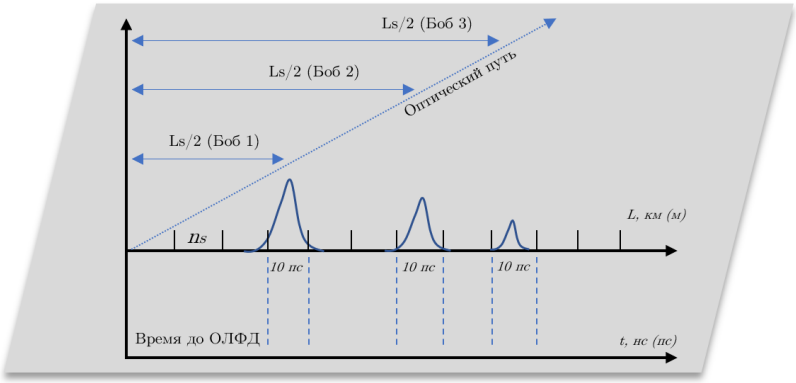



*Рис. 2. Пространственно-временной поиск оптического сигнала*

Обнаружение оптического синхросигнала осуществляется последовательным анализом временных интервалов *ns.* После каждой посылки импульса ОЛФД станции Алиса настраиваются на приём сигнала через заданные временные интервалы. Результаты срабатываний фиксируются. Детально процесс обнаружения для топологии «точка-точка» описан в работах [8, 9]. В нашем случае (рис. 1) результатом синхронизации будут считаться обнаруженные сигнальные интервалы и их интерпретация во временные данные. Так, станция Алиса будет знать в какой момент времени необходимо активировать ОЛФД для каждого сегмента (С1, С2 и С3 на рис. 1), а электроника станций Боб 1, Боб 2, Боб 3 будет обладать информацией в какой момент времени прикладывать напряжение на ФМ. Отличительной особенностью описываемой схемы является наличие трех вре-

менных интервалов и трех пространственных отрезков, соответствующих расстояниям до сегментов Боб 1 – Боб 3. *Технически мы не видим препятствий использования одной базовой станции (одного ЛД и двух ОЛФД) для осуществления синхронизации с тремя удаленными станциями. Естественно, с точки зрения работы квантового протокола, все станции должны быть связаны аутентичным общедоступным каналом, но для процесса калибровки этого не требуется.*

Проведем усредненный анализ потерь оптического сигнала, не учитывающий возможные дестабилизирующие факторы, способные повлиять на КК. Пусть расстояние от станции *C0* до *C1* равно *10 км*, от *C1* до *C2 – 15 км*, от *C2* до *C3* составляет *20 км*. Таким образом, максимальное расстояние от источника излучения до ЗФ *C3* и обратно до ОЛФД *C0* равно *90 км*. Собственные потери в КК для одномодового волокна и потери на разъемных соединениях *(lf)* принимаем равными *0.2 дБ/км.* Потери на сварных соединениях *(lw) – 0.05 дБ*. Суммарные потери в сегменте *lc – 10 дБ.*

$$L\text{сумм для С1} = 0.2[lf \text{ в ВОЛС}] * 10 + 0.2[lf] * 8 + 0.05\,[lw] * 28 + 10[lc] = 15 \text{ дБ}$$

$$L\text{сумм для С2} = 0.2[lf \text{ в ВОЛС}] * 25 + 0.2[lf] * 8 + 0.05\,[lw] * 62 + 10[lc] = 19.7 \text{ дБ}$$

$$L\text{сумм для С3} = 0.2[lf \text{ в ВОЛС}] * 45 + 0.2[lf] * 8 + 0.05\,[lw] * 98 + 10[lc] = 25.5 \text{ дБ}$$

Следует обратить внимание на то, что приведенные потери справедливы только для прямого направления, т.е. при построении энергетической модели системы КРК и расчете мощности оптического импульса, вносимого аттенюатором затухания, необходимо учитывать обратный путь от каждого сегмента.

**Выводы и дискуссия.** В статье рассмотрен способ обнаружения оптического сигнала синхронизации для участка сети квантовых коммуникаций, особенностью которого является наличие одной приемо-передающей станции и нескольких кодирующих станций. Описан принцип работы системы квантового распределения ключей с фазовым кодированием и предложена концепция алгоритма, адаптированного для предлагаемой конфигурации городской квантовой сети, содержащей несколько последовательных сегментов. Приведен расчет усредненных потерь в квантовом канале для наглядного понимания вносимых затуханий.

Переходя к дискуссии, можно выделить несколько актуальных проблем по мнению автора при технической реализации квантовых сетей: защищенность каналов аутентификации (как обеспечить безусловную защищенность не только квантового протокола, но и процесса аутентификации? Можно ли обойтись без классической криптографии при первичной аутентификации? Насколько безопасно использовать системы КРК, если злоумышленник имеет доступ к каналу синхронизации, аутентификации?) [11–16]; безопасная реализация самих ДПУ (вероятность НСД к аппаратуре в ДПУ больше вероятности атак на квантовый канал между ДПУ? Как обеспечить доставку квантовых ключей конечному пользователю?); однофотонность при формировании ключей (насколько реально добиться однофотонной передачи на расстоянии, например, 40 км в городских условиях? Какой протокол с доказанной стойкостью можно использовать в реальных условиях эксплуатации?) [17–20].

*Автор статьи благодарен читателю и приглашает дать обратную связь по приведенным вопросам.*

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Chen Y.A. et al.* An integrated space-to-ground quantum communication network over 4,600 kilometres // Nature. – 2021. – Vol. 589, No. 7841. – P. 214-219.
2. *Gisin N., Ribordy G., Tittel W., Zbinden H.* Quantum cryptography // Reviews of Modern Physics. – 2002. – Vol. 74, No. 1. – P. 145-195.
3. *Bennett C.H., Brassard G., & Ekert A.K.* Quantum Cryptography // Scientific American. – 1992. – 267 (4). – P. 50-57. – http://www.jstor.org/stable/24939253.
4. *Кулик С.* Квантовая криптография // Фотоника. – 2010. – №. 2. – С. 36-41.
5. *Кулик С.П., Молотков С.Н.* MDI–Measurement Device Independent квантового распределения ключей // Письма в Журнал экспериментальной и теоретической физики. – 2023. – Т. 118, № 1. – С. 62-70.

6. *Lo H.K., Curty M., Qi B.* Measurement-device-independent quantum key distribution // Physical review letters. – 2012. – Vol. 108, No. 13. – P. 130503.
7. *Bouwmeester D. et al.* Experimental quantum teleportation // Nature. – 1997. – Vol. 390, No. 6660. – P. 575-579.
8. *Pljonkin A., Rumyantsev, K., Singh, P.K.* Synchronization in quantum key distribution systems // Cryptography. – 2017, 1, 18. – DOI: 10.3390/cryptography1030018.
9. *Румянцев К.Е., Плёнкин А.П.* Синхронизация системы квантового распределения ключа при использовании фотонных импульсов для повышения защищённости // Известия ЮФУ. Технические науки. – 2014. – № 8. – С. 81-96.
10. *Гальярди Р.М., Карп Ш.* Оптическая связь: пер. с англ. / под ред. А.Г. Шереметьева. – М.: Связь, 1978. – 424 с.
11. *Deng F.G., & Long G.L.* Secure direct communication with a quantum one-time pad // Physical Review A. – 2004. – 69 (5). – 052319.
12. *Pljonkin A., Petrov D., Sabantina L., Dakhkilgova K*. Nonclassical attack on a quantum keydistribution system // Entropy. – 2021. – Vol. 23, No. 5. – DOI: 10.3390/e23050509.
13. *Zhao Y., Fung C.H.F., Qi B., Chen C., & Lo H.K*. Quantum hacking: Experimental demonstration of time-shift attack against practical quantum-key-distribution systems // Physical Review A. – 2008. – 78 (4). – 042333.
14. *Makarov V., & Hjelme D.R.* Faked states attack on quantum cryptosystems // Journal of Modern Optics. – 2005. – 52 (5). – P. 691-705.
15. *Сабанов А.Г., Шелупанов А.А.* Идентификация и аутентификация в цифровом мире. – М.: Горячая Линия–Телеком, 2022.
16. *Кравцов К.С. и др.* Система релятивистской квантовой криптографии. – 2018.
17. *Beals T.R., Sanders B.C*. Distributed relay protocol for probabilistic information-theoretic security in a randomly-compromised network // Information Theoretic Security: Third International Conference, ICITS 2008, Calgary, Canada, August 10-13, 2008. Proceedings 3. – Springer Berlin Heidelberg, 2008. – P. 29-39.
18. *Dianati M., Alléaume R.* Architecture of the Secoqc quantum key distribution network // 2007 First International Conference on Quantum, Nano, and Micro Technologies (ICQNM'07). – IEEE, 2007. – P. 13-13.
19. *Barnett S.M., Phoenix S.J.* D. Securing a quantum key distribution relay network using secret sharing // 2011 IEEE GCC Conference and Exhibition (GCC). – IEEE, 2011. – P. 143-145.
20. *Поздняков А.М.* Способ передачи сообщения через вычислительную сеть с применением аппаратуры квантового распределения ключей. – 2019.

REFERENCES

1. *Chen Y.A. et al.* An integrated space-to-ground quantum communication network over 4,600 kilometres, *Nature,* 2021, Vol. 589, No. 7841, pp. 214-219.
2. *Gisin N., Ribordy G., Tittel W., Zbinden H.* Quantum cryptography, *Reviews of Modern Physics,* 2002, Vol. 74, No. 1, pp. 145-195.
3. *Bennett C.H., Brassard G., & Ekert A.K.* Quantum Cryptography, *Scientific American,* 1992, 267 (4), pp. 50-57. Available at: http://www.jstor.org/stable/24939253.
4. *Kulik S.* Kvantovaya kriptografiya [Quantum cryptography], *Fotonika* [Photonics], 2010, No. 2, pp. 36-41.
5. *Kulik S.P., Molotkov S.N.* MDI–Measurement Device Independent kvantovogo raspredeleniya klyuchey [MDI-Measurement Device Independent of quantum key distribution], *Pis'ma v Zhurnal eksperimental'noy i teoreticheskoy fiziki* [Letters to the Journal of Experimental and Theoretical Physics], 2023, Vol. 118, No. 1, pp. 62-70.
6. *Lo H.K., Curty M., Qi B.* Measurement-device-independent quantum key distribution, *Physical review letters,* 2012, Vol. 108, No. 13, pp. 130503.
7. *Bouwmeester D. et al.* Experimental quantum teleportation, *Nature,* 1997, Vol. 390, No. 6660, pp. 575-579.
8. *Pljonkin A., Rumyantsev, K., Singh, P.K.* Synchronization in quantum key distribution systems, *Cryptography,* 2017, 1, 18. DOI: 10.3390/cryptography1030018.
9. *Rumyantsev K.E., Plenkin A.P.* Sinkhronizatsiya sistemy kvantovogo raspredeleniya klyucha pri ispol'zovanii fotonnykh impul'sov dlya povysheniya zashchishchennosti [Synchronization of the quantum key distribution system using photon pulses to increase security], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2014, No. 8, pp. 81-96.
10. *Gal'yardi R.M., Karp Sh.* Opticheskaya svyaz' [Optical communication]: trans. from engl., ed. by A.G. Sheremet'eva. Moscow: Svyaz', 1978, 424 p.

11. *Deng F.G., & Long G.L.* Secure direct communication with a quantum one-time pad, *Physical Review A,* 2004, 69 (5), 052319.
12. *Pljonkin A., Petrov D., Sabantina L., Dakhkilgova K*. Nonclassical attack on a quantum keydistribution system, *Entropy,* 2021, Vol. 23, No. 5. DOI: 10.3390/e23050509.
13. *Zhao Y., Fung C.H.F., Qi B., Chen C., & Lo H.K*. Quantum hacking: Experimental demonstration of time-shift attack against practical quantum-key-distribution systems, *Physical Review A,* 2008, 78 (4), 042333.
14. *Makarov V., & Hjelme D.R.* Faked states attack on quantum cryptosystems, *Journal of Modern Optics,* 2005, 52 (5), pp. 691-705.
15. *Sabanov A.G., Shelupanov A.A.* Identifikatsiya i autentifikatsiya v tsifrovom mire [Identification and authentication in the digital world]. Moscow: Goryachaya Liniya–Telekom, 2022.
16. *Kravtsov K.S. i dr.* Sistema relyativistskoy kvantovoy kriptografii [The system of relativistic quantum cryptography], 2018.
17. *Beals T.R., Sanders B.C*. Distributed relay protocol for probabilistic information-theoretic security in a randomly-compromised network, *Information Theoretic Security: Third International Conference, ICITS 2008, Calgary, Canada, August 10-13, 2008. Proceedings 3*. Springer Berlin Heidelberg, 2008, pp. 29-39.
18. *Dianati M., Alléaume R.* Architecture of the Secoqc quantum key distribution network, *2007 First International Conference on Quantum, Nano, and Micro Technologies (ICQNM'07)*. IEEE, 2007, pp. 13-13.
19. *Barnett S.M., Phoenix S.J.* D. Securing a quantum key distribution relay network using secret sharing, *2011 IEEE GCC Conference and Exhibition (GCC)*. IEEE, 2011, pp. 143-145.
20. *Pozdnyakov A.M.* Sposob peredachi soobshcheniya cherez vychislitel'nuyu set' s primeneniem apparatury kvantovogo raspredeleniya klyuchey [Method for transmitting messages through a computing network using quantum key distribution equipment], 2019.

Статью рекомендовал к опубликованию д.т.н., профессор И.И. Турулин.

**Плёнкин Антон Павлович** – Южный федеральный университет; e-mail: pljonkin@sfedu.ru; г. Таганрог, Россия; тел.: 89054592158; кафедра ИБТКС; к.т.н.; доцент.

**Pljonkin Anton Pavlovich** – Southern Federal University; e-mail: pljonkin@sfedu.ru; Taganrog, Russia; phone: +79054592158; the Department of Information Security of Telecommunication Systems; cand. of eng. sc.; associate professor.

УДК 621.396.67 DOI 10.18522/2311-3103-2024-5-260-270

**А.В. Геворкян, В.С. Савостин**

# СВЕРХШИРОКОПОЛОСНЫЕ РЕШЁТКИ АНТЕНН ВИВАЛЬДИ С ТЕМ-РУПОРОМ

*Приведены конструкции и характеристики антенных решёток на основе антиподного излучателя Вивальди. Исследуются антенные решётки с ТЕМ-рупорами линейного и эллиптического профиля. Проведена оптимизация параметров рупоров. Характеристики исследовались в диапазоне частот от 4 до 12 ГГц. Антенная решётка с ТЕМ-рупором линейного профиля имеет лучший КСВН в диапазоне от 4 до 5 ГГц (для крайних излучателей максимум равен 4,75, а для остальных – 3,33). Рабочая полоса частот антенной решётки находится в диапазоне от 4,90 до 12,00 ГГц (коэффициент перекрытия $k_п$=2,45). Частотная характеристика реализованного коэффициента усиления (КУ) имеет провалы. Антенная решётка с ТЕМ-рупор эллиптического профиля с узким основанием имеет минимальную рабочую полосу частот (от 7,06 до 12,00 ГГц ($k_п$=1,70)) и плавную характеристику реализованного КУ. Антенная решётка с увеличенной шириной основания ТЕМ-рупора эллиптического профиля имеет лучший КСВН в диапазоне от 5,3 до 12,0 ГГц (для крайних излучателей максимум равен 2,51, а для остальных – 2,15), но характеристика реализованного КУ плавная только до 9 ГГц. Рабочая полоса частот антенной решётки находится в диапазоне от 4,84 до 12,00 ГГц ($k_п$=2,48). Лучшие характеристики у антенной решётки с ТЕМ-рупором эллиптического профиля с расширенным основанием и увеличенной высотой. Увеличение высоты рупора приводит к увеличению значений реализованного КУ на частотах более 9,25 ГГц, где были провалы. Рабочая полоса частот находится в диапазоне от 4,72 до 12,00 ГГц*

*($k_п$=2,54). В рабочей полосе частот значения реализованного КУ находятся в диапазоне от 11,9 до 20,6 дБ. Таким образом, выбором формы и параметров рупора можно улучшить частотные характеристики антенной решётки.*

*Антенная решётка; антиподная антенна Вивальди; ТЕМ-рупор эллиптического профиля; сверхширокополосность; КСВН; реализованный коэффициент усиления.*

**A.V. Gevorkyan, V.S. Savostin**

**ULTRA-WIDEBAND VIVALDI ANTENNA ARRAYS WITH TEM HORN**

*Presents the designs and characteristics of antenna array based on the antipodal Vivaldi element. Antenna arrays with TEM horns of linear and elliptical profile are studied. The parameters of the horns were optimized. The characteristics were studied in the frequency range from 4 to 12 GHz. The antenna array with a linear profile TEM horn has the best VSWR in the range from 4 to 5 GHz (for the edge elements the maximum is 4,75, and 3,33 for others). The operating frequency band of the antenna array is in the range from 4,90 to 12,00 GHz (overlap coefficient $k_п$=2,45). The frequency characteristics of the realized gain has dips. The antenna array with a TEM horn of an elliptical profile with a narrow base has a minimum operating frequency band (from 7,06 to 12,00 GHz ($k_п$=1,70) and a smooth characteristic of the realized gain. Antenna array with an increased base width of the TEM horn of an elliptical profile has the best VSWR in the range from 5,3 to 12,0 GHz (for edge elements, the maximum is 2,51, and 2,15 for others), but the characteristics of the realized gain is smooth only up to 9 GHz. The operating frequency band of the antenna array is in the range from 4,84 to 12,00 GHz ($k_п$=2,48). The best characteristics has the antenna array with a TEM horn elliptical profile with an expanded base and increased height. An increase in the height of the horn leads to an increase in the values of the realized gain at frequencies above 9,25 GHz, where there were dips. The operating frequency band ranges from 4,72 to 12,00 GHz ($k_п$=2,54). In the operating frequency band, the values of the realized gain are in the range from 11,9 to 20,6 dB. Thus, by choosing the shape and parameters of the horn, the frequency characteristics of the antenna array can be improved.*

*Antenna array; antipodal Vivaldi antenna; elliptical profile TEM-horn; ultra-wideband; VSWR; realized gain.*

**Введение.** Широкое применение сверхширокополосных (СШП) антенн вызвано активным развитием и внедрением сверхширокополосных систем. Эти антенны обеспечивают требуемые характеристики в сверхшироком диапазоне рабочих частот (коэффициент перекрытия таких антенн $k_{\text{п}} = \frac{f_{max}}{f_{min}} \geq 2$, где $f_{max}$ и $f_{min}$ – максимальная и минимальная рабочая частота антенны по заданному уровню КСВН).

Одной из самых широко используемых разновидностей СШП антенн [1] является антенна Вивальди. Широкое применение антенн Вивальди вызвано их преимуществами [2] по сравнению с другими типами СШП антенн: сверхширокополосность, большой коэффициент усиления (КУ) одиночных антенн и простота изготовления. Такая антенна впервые предложена Гибсоном [3] в 1979 году и представляла собой проводник с расширяющейся щелью, который расположен на диэлектрической подложке.

Изначально предложена компланарная конструкция антенны Вивальди. Такая антенна имеет большой недостаток – её полоса рабочих частот ограничена структурой линии питания. Для расширения рабочей полосы частот, Газитом [4] в 1988 году предложен антиподный тип антенны Вивальди. В ней излучающие проводники находятся на противоположных сторонах подложки.

Для улучшения кросс-поляризационных свойств, в 1993 году Лэнгли [5] предложена новая разновидность антенны Вивальди – балансная антиподная антенна Вивальди. Антенна имеет две подложки и три проводника (два внешних проводника подключаются к оплётке коаксиального кабеля, а средний проводник подключается к жиле).

В дальнейшем, интерес к антенне Вивальди вырос, и изменение её конструкции продолжается до сих пор [6–8].

Данная статья посвящена разработке линейных 10-элементных эквидистантных антенных решёток Вивальди. Конструкция используемого антиподного излучателя Вивальди основана на одиночной антенне [9], которая оптимизирована для работы в составе

антенной решётки [10]. Основной особенностью разработанных антенных решёток Вивальди является их совмещение с ТЕМ-рупором линейного [11–17] и эллиптического [18–22] профиля.

**Конструкция излучателя.** На рис. 1 приведена конструкция антиподного излучателя Вивальди, смоделированного в HFSS [23]. Он состоит из двух проводников, расположенных на диэлектрической подложке из материала FR-4. Питание обеспечивает 50-Омный коаксиальный кабель. Параметры излучателя приведены в таблице.

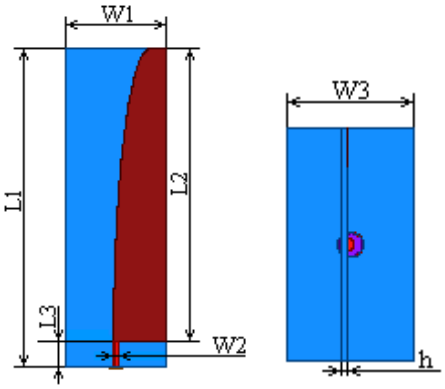



*Рис. 1. Конструкция излучателя Вивальди: а – вид спереди, б – вид сверху*

Таблица

**Параметры антиподного излучателя Вивальди**

| Конструктивный параметр | Значение параметра, мм |
|---|---|
| Ширина антенны – W1 | 18,5 |
| Длина антенны – L1 | 58 |
| Длина проводников – L2 | 53,5 |
| Длина линии питания – L3 | 4,5 |
| Ширина линии питания – W2 | 0,4 |
| Ширина экрана – W3 | 10 |
| Толщина подложки – h | 1 |

**Конструкции антенных решёток**

***А. Антенная решётка с ТЕМ-рупором линейного профиля.*** Первая конструкция антенной решётки приведена на рис. 2. Она представляет собой линейную 10-элементную эквидистантную антенную решётку с параллельным питанием. Антенная решётка совмещена с ТЕМ-рупором линейного профиля. Параметры рупора: ширина основания – 10 мм, высота – 70 мм и ширина раскрыва – 90 мм.

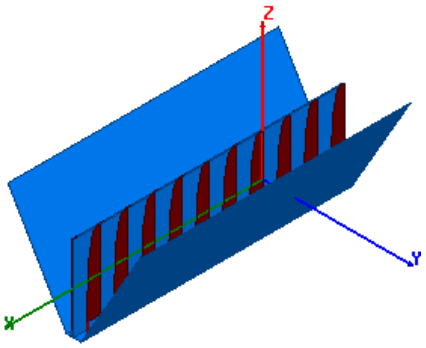



*Рис. 2. Конструкция антенной решётки*

***Б. Антенная решётка с ТЕМ-рупором эллиптического профиля.*** Особенностью следующей антенной решётки является изменение формы рупора – ТЕМ-рупор имеет эллиптический профиль. На рис. 3 приведена конструкция антенной решётки. Параметры ТЕМ-рупора: ширина основания – 10 мм, высота – 75 мм и ширина раскрыва – 60 мм.

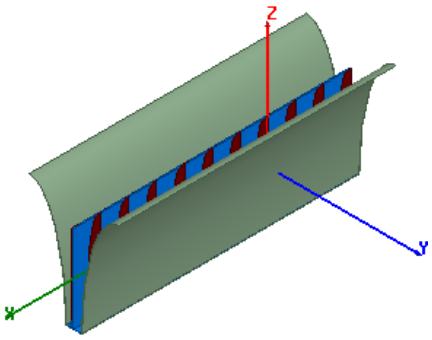



*Рис. 3. Конструкция антенной решётки*

***В. Антенная решётка с ТЕМ-рупором эллиптического профиля и расширенным основанием.*** Для уменьшения влияния стенок рупора на согласование антенн с линией питания, ширина основания рупора была увеличена до 30 мм (ширина раскрыва – 80 мм). Полученная конструкция приведена на рис. 4.

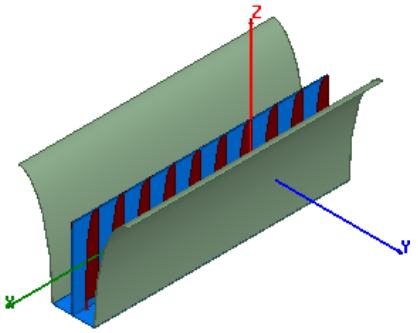



*Рис. 4. Конструкция антенной решётки*

***Г. Антенная решётка с ТЕМ-рупором эллиптического профиля, расширенным основанием и увеличенной высотой.*** Для улучшения частотной характеристики реализованного КУ, высота стенок рупора была увеличена до 85 мм. Остальные параметры остались без изменений. Полученная конструкция приведена на рис. 5

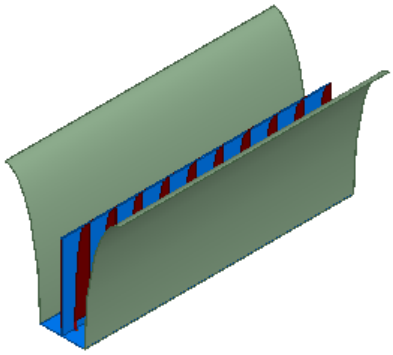


*Рис. 5. Конструкция антенной решётки*

**Результаты моделирования характеристик антенных решёток**

***А. Антенная решётка с ТЕМ-рупором линейного профиля.*** На рис. 6 приведены частотные характеристики КСВН, реализованного КУ и КПД антенной решётки. Здесь и далее, на рисунках с КСВН, штриховые и пунктирные линии соответствуют крайним (первому и последнему) излучателям антенной решётки. Для них максимальное значение КСВН принимается равным 4, а для остальных – 3.

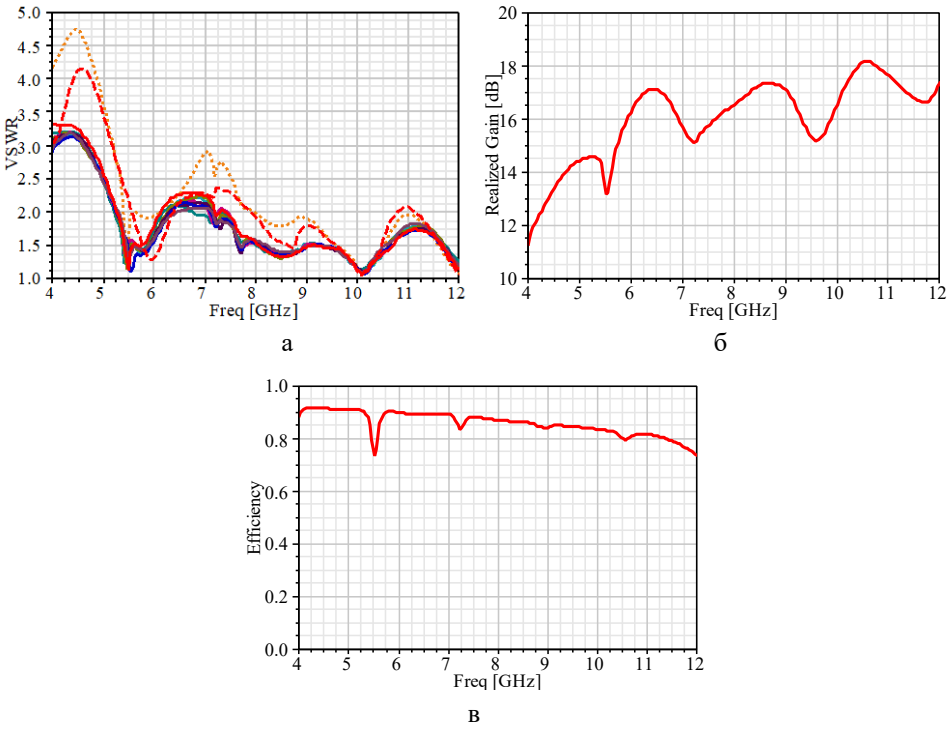



*Рис. 6. Частотные характеристики: а – КСВН, б – реализованного КУ, в – КПД*

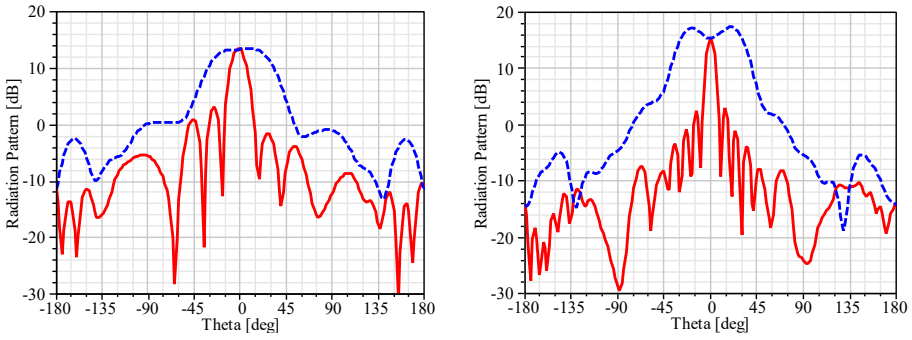
Частотная характеристика КСВН показывает, что рабочая полоса частот антенной решётки находится в диапазоне от 4,90 до 12,00 ГГц ($k_п$=2,45). В диапазоне от 4 до 5 ГГц эта антенная решётка имеет лучший КСВН.

Частотная характеристика реализованного КУ (Theta=0°) показывает, что его значения находятся в диапазоне от 13,2 до 18,2 дБ (рис. 6,б). Характеристика имеет провалы. Для понимания причины провалов в частотной характеристике, на рис. 7 приведены диаграммы направленности (ДН) для Е- ( — ) и Н-плоскостей ( - - ). Из рисунка видно, что провалы вызваны расширением ДН (из-за того, что распределение поля в раскрыве рупора не является равномерным). Поэтому, для улучшения распределения поля, в следующей конструкции ТЕМ-рупора линейного профиля был заменен на ТЕМ-рупор эллиптического профиля.

Частотная характеристика КПД (см. рис. 6,в) показывает, что он превышает 73%. Уменьшение КПД с ростом частоты связано с ростом тепловых потерь в диэлектрической подложке.

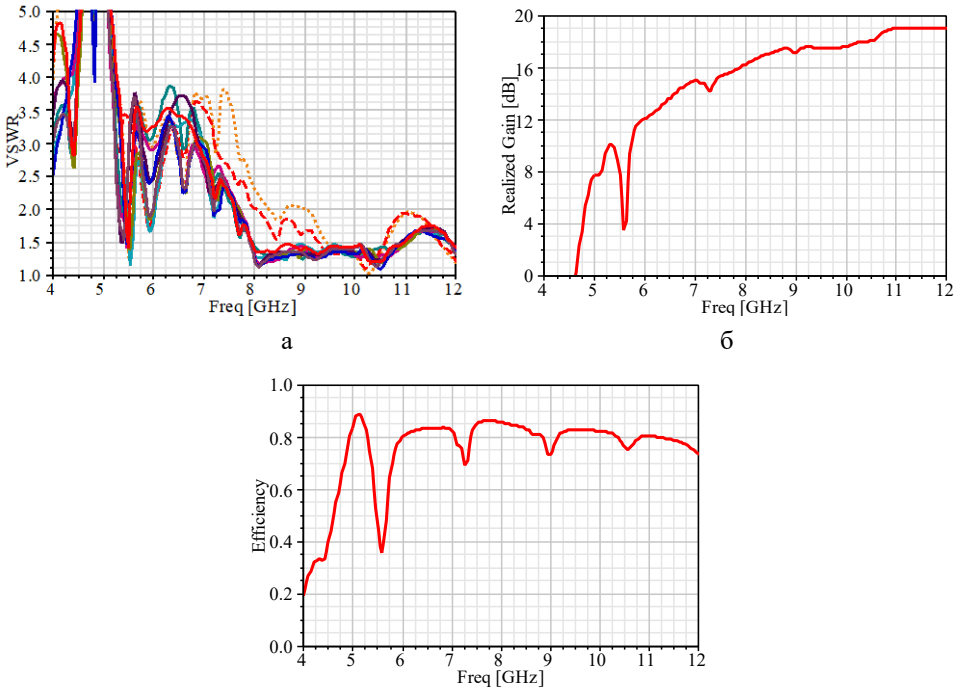
***Б. Антенная решётка с ТЕМ рупором эллиптического профиля.*** На рис. 8 приведены частотные характеристики КСВН, реализованного КУ и КПД антенной решётки.

Частотная характеристика КСВН (см. рис. 8,а) показывает, что ТЕМ-рупор эллиптического профиля вызывает сильное увеличение КСВН на низких частотах. На высоких частотах значения КСВН уменьшаются. Рабочая полоса частот находится в диапазоне от 7,06 до 12,00 ГГц ($k_п$=1,70). Изменение значений КСВН связано с близким расположением нижней части стенок рупора к излучателям.




а б

*Рис. 7. ДН антенной решётки на частотах: а – 5,56 ГГц, б – 9,60 ГГц*




а б

в

*Рис. 8. Частотные характеристики: а – КСВН, б – реализованного КУ, в – КПД*

Сравнение рис. 6,б и 8,б показывает, что на высоких частотах происходит увеличение реализованного КУ (в рабочей полосе его значения находятся в диапазоне от 14,2 до 19,0 дБ). Характеристика имеет плавный вид. Однако в районе частоты 5,6 ГГц имеется большой провал. Из рис. 9 видно, что на этой частоте, в направлении Theta=0°, ДН имеет провал, а два её максимума отличаются по значению.

Из частотной характеристики КПД (см. рис. 8,в) видно, что провалы также увеличились. Наибольший из них наблюдается на частоте 5,6 ГГц.

Провалы в частотных характеристиках реализованного КУ и КПД связаны с резонансными явлениями, свойственные рассматриваемого излучателю, и параметры рупора влияют на них.

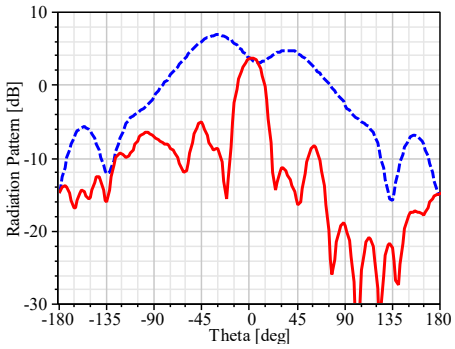



*Рис. 9. ДН антенной решётки на частоте 5,6 ГГц*

***В. Антенная решётка с ТЕМ-рупором эллиптического профиля и расширенным основанием.*** На рис. 10 приведены частотные характеристики КСВН, реализованного КУ и КПД антенной решётки. Видно, что увеличение ширины основания рупора улучшило частотные характеристики КСВН и КПД. Рабочая полоса частот антенной решётки находится в диапазоне от 4,84 до 12,00 ГГц ($k_п$=2,48). Из всех антенных решёток, она имеет лучший КСВН в диапазоне от 5,3 до 12 ГГц (для крайних излучателей максимум равен 2,51, а для остальных – 2,15). В рабочей полосе частот значения КПД превышают 78%.

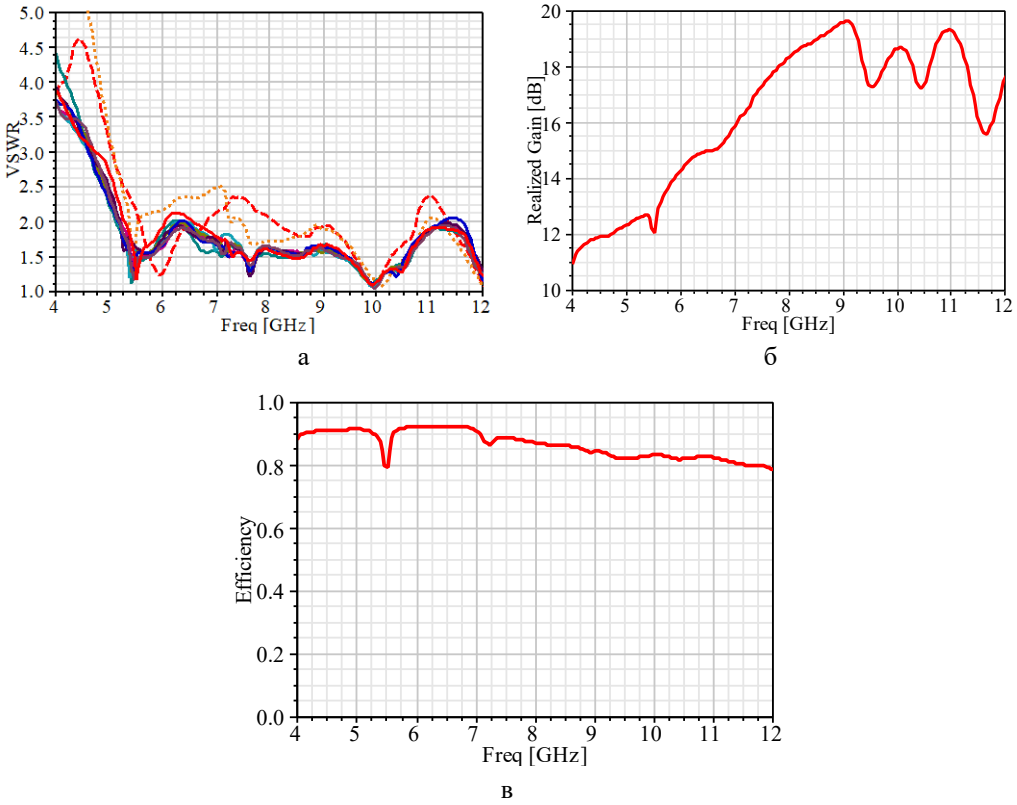



*Рис. 10. Частотные характеристики: а – КСВН, б – реализованного КУ, в – КПД*

Расширение основания рупора, как следует из сравнения рис. 8,б и 10,б приводит к существенному уменьшению провала частотной характеристики реализованного КУ в районе 5,5–5,6 ГГц и увеличению его значений на частотах до 9 ГГц. Они находятся в

диапазоне от 12,1 до 19,7 дБ. В остальной части исследованного диапазона частот характеристика ухудшается и имеет провалы. Причина ухудшений, как и в рассмотренных выше случаях, заключается в расширении ДН в Н-плоскости.

***Г. Антенная решётка с ТЕМ-рупором эллиптического профиля, расширенным основанием и увеличенной высотой.*** На рис. 11 приведены частотные характеристики КСВН, реализованного КУ и КПД.

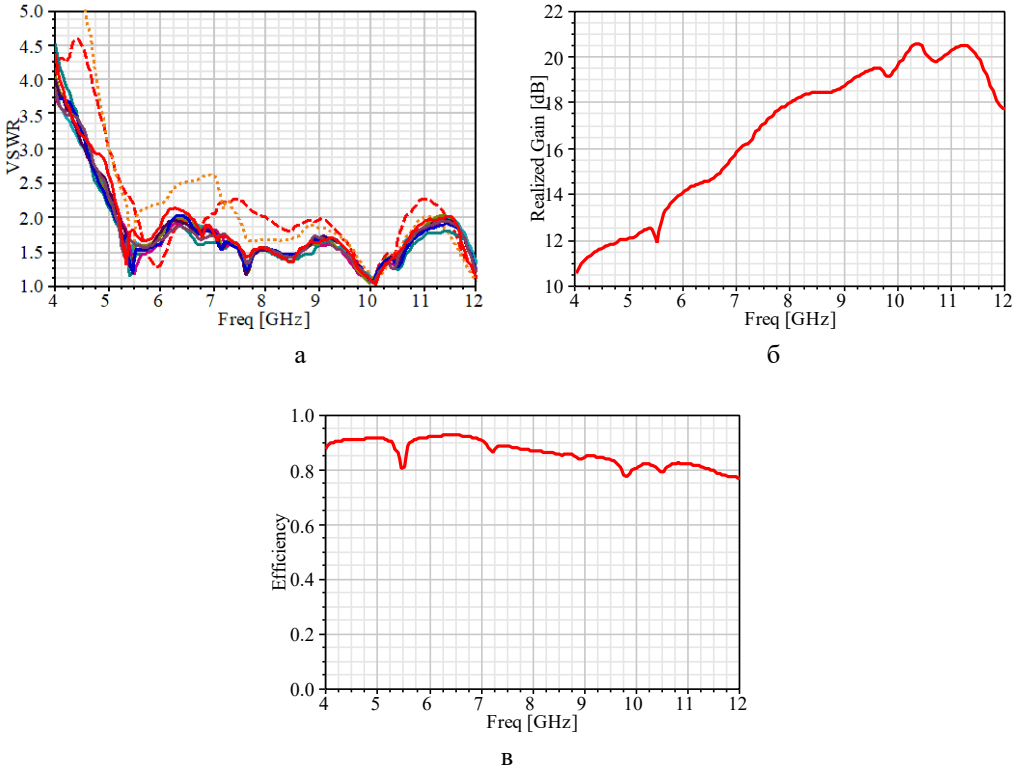
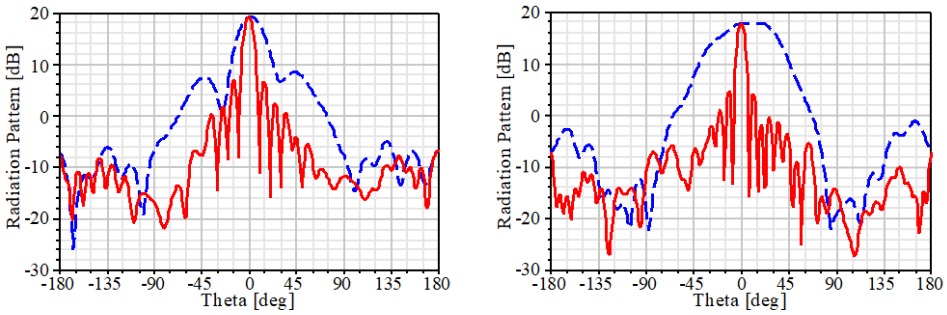



*Рис. 11. Частотные характеристики: а – КСВН, б – реализованного КУ, в – КПД*

Частотная характеристика КСВН (см. рис. 11,а) показывает, что его значение у крайних излучателей превышает 4 в диапазоне 4 – 4,8 ГГц. КСВН остальных излучателей не превышает 3, начиная с частоты 4,72 ГГц. Следовательно, рабочая полоса частот антенной решётки находится в диапазоне от 4,72 до 12,00 ГГц ($k_п$=2,54).

Рис. 11,б показывает, что увеличение высоты рупора приводит к увеличению на 0,8–3,8 дБ значений реализованного КУ на частотах более 9,25 ГГц и провалы в характеристике уменьшились. На остальных частотах он не изменяется или уменьшается (максимум – 0,5–0,9 дБ в диапазоне 8,5–9,2 ГГц). Улучшение связано с тем, что распределение поля в раскрыве рупора становится равномернее. В рабочей полосе частот значения реализованного КУ находятся в диапазоне от 11,9 до 20,6 дБ, а значения КПД (см. рис. 11,в) превышают 76%.

На рис. 12 приведены ДН в Е- ( ——— ) и Н-плоскостях ( – – – ). Из рисунка видно, что на частоте 12 ГГц в Н-плоскости ДН шире, чем на меньшей частоте 9,8 ГГц. Поэтому провал в районе частоты 12 ГГц остается. Это может быть связано и с тем, что на этой частоте электрическая ширина излучателя равна 0,74 λ.




а б

*Рис. 12. ДН антенной решётки на частотах: а – 9,8 ГГц, б – 12,0 ГГц*

**Заключение.** Результаты исследования показывают, что выбором формы и параметров рупора можно улучшить частотные характеристики антенной решётки. Рабочая полоса частот первого варианта конструкции антенной решётки с ТЕМ-рупором линейного профиля находится в диапазоне от 4,90 до 12,00 ГГц (коэффициент перекрытия $k_п$=2,45), но частотная характеристика реализованного КУ имеет провалы, а его значения находятся в диапазоне от 13,2 до 18,2 дБ. Последняя конструкция является лучшей с точки зрения улучшения характеристик в пределах диапазона рабочих частот. Она имеет большую ширину рабочей полосы частот (в диапазоне от 4,72 до 12,00 ГГц ($k_п$=2,54)), частотная характеристика реализованного КУ имеет плавный вид, а его значения находятся в диапазоне от 11,9 до 20,6 дБ. Особенности последней конструкции: ТЕМ-рупор имеет эллиптический профиль, увеличенную ширину основания (30 мм) и высоту (85 мм).

*Работа выполнена при финансовой поддержке Российского научного фонда (Проект № 22-19-00537, https://rscf.ru/project/22-19-00537/) в Центре коллективного пользования «Прикладная электродинамика и антенные измерения» Южного федерального университета, г. Таганрог.*

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Tayebi M., Dastranj A.A., Alighanbari A.* Ultra Wide Band Antipodal Vivaldi Antenna with Tapered Triangular Corrugated Edges // IEEE 27th Iranian Conference on Electrical Engineering (ICEE). – 2019. – P. 1637-1642. – DOI: 10.1109/IranianCEE.2019.8786722.
2. *Tangwachirapan S., Thaiwirot W., Akkaraekthalin P.* Design of Ultra-Wideband Antipodal Vivaldi Antenna with Square Dielectric Lens for Microwave Imaging Applications // IEEE 7th International Electrical Engineering Congress (iEECON). – 2019. – P. 1-4. – DOI: 10.1109/iEECON45304.2019.8939032.
3. *Gibson P.J.* The Vivaldi Aerial // IEEE 9th European Microwave Conference. – 1979. – P. 101-105. – DOI: 10.1109/EUMA.1979.332681.
4. *Gazit E.* Improved Design of the Vivaldi Antenna. Microwaves // IEE Proceedings. – 1988. – Vol. 135. – Pt. H. N. 2. – P. 89-92. – DOI: 10.1049/ip-h-2.1988.0020.
5. *Langley J.D.S., Hall P.S., Newham P.* Novel ultrawide-bandwidth Vivaldi antenna and low crosspolarisation // Electronics Letters. – 1993. – Vol. 29, Issue 23. – P. 2004-2005. – DOI: 10.1049/el:19931336.
6. *Logan J.T., Kindt R.W., Vouvakis M.N.* A 1.2–12 GHz Sliced Notch Antenna Array // IEEE Transactions on Antennas and Propagation. – Apr. 2018. – Vol. 66, No. 4. – P. 1818-1826. – DOI: 10.1109/TAP.2018.2809476.
7. *Ning Y., Ling L., Bao F., Li Z.* An Ultra-wideband Miniaturized Antipodal Vivaldi Antenna // 2018 International Conference on Microwave and Millimeter Wave Technology ICMMT). – 2018. – P. 1-3. – DOI: 10.1109/ICACCI.2018.8563335.
8. *Minjie Guo et al.* High-gain antipodal Vivaldi antenna with metamaterial covers // IET Microwaves, Antennas & Propagation. – 2019. – Vol. 13, Iss. 15. – DOI: 10.1049/iet-map.2019.0449.
9. *Kumar R., Behera B. R., Suraj P.* A Modified Leaf Shaped Antipodal Vivaldi Antenna for UWB Applications // 2018 IEEE Indian Conference on Antennas and Propogation (InCAP). – 2018. – P. 1-4. – DOI: 10.1109/INCAP.2018.8770788.

10. *Suo Y., Qi F., Li W.* Design of ultra-wideband TEM horn antenna for life detection // 2021 IEEE International Symposium on Antenna and Propagation (ISAP). – 2021. – P. 1–2. – DOI: 10.23919/ISAP47258.2021.9614515.
11. *Wang D., Zheng S., Deng F., Hou D.* Analysis on the time-domain transfer characteristics of UWB horn antenna // IEEE 5th International Symposium on Microwave, Antenna, Propagation and EMC Technologies for Wireless Communications. – 2013. – P. 351-354. – DOI: 10.1109/MAPE.2013.6689819.
12. *Wang Y., Chen Y.-G.* Research on the impact of TEM horn antenna's structural parameter pulse waveform // 2011 IEEE International Conference on Electrical and Control Engineering. – 2011. – P. 3155-3138. – DOI: 10.1109/ICECENG.2011.6057399.
13. *Zalabsky T., Bezousek P.* TEM horn antenna for high energy emission // IEEE 23rd International Conference Radioelektronika (RADIOELEKTRONIKA). – 2013. – P. 92-95. – DOI: 10.1109/RadioElek.2013.6530898.
14. *Lin S., Yu S., Jiao J.-L., Yang C.-T.* Simulation and analysis of an ultra-wideband TEM horn antenna with edge // 2017 IEEE International Symposium on Antennas and Propagation (ISAP). – 2017. – P. 1-2. – DOI: 10.1109/ISANP.2017.8228996.
15. *Schoeman K., Meyer P., D.I.L. de Villiers.* Exponential TEM horn with a convex triangular arc // 2013 IEEE Africon. – 2013. – P. 1-5. – DOI: 10.1109/AFRCON.2013.6757614.
16. *Калошин В.А., Нгуен К.З.* Исследование характеристик СШП плоских двумерно-периодических решеток ТЕМ рупоров // Журнал радиоэлектроники. – 2017. – № 5. – http://jre.cplire.ru/jre/may17/14/text.pdf.
17. *Калошин В.А., Ле Н.Т., Фролова Е.В.* Сверхдиапазонная цилиндрическая антенная решетка ТЕМ рупоров // Журнал радиоэлектроники. – 2020. – № 4. – https://doi.org/10.30898/1684-1719.2020.4.2.
18. *Suo Y., Qi F., Li W.* Design of exponential gradient TEM horn antenna for ground penetrating radar // 2021 IEEE International Symposium on Antennas and Propagation (ISAP). – 2021. – P. 1-2. – DOI: 10.23919/ISAP47258.2021.9614604.
19. *Li Z.-H., Ma J., Wu J., Huang J.-J., Lu Y. -H., Peng L.* Miniaturization Design of TEM Horn Antenna for Ground-Penetrating Radar // 2023 IEEE International Conference on Microwave and Millimeter Wave Technology (ICMMT). – 2023. – P. 1-3. – DOI: 10.1109/ICACCI.2018.8554597.
20. *Ameri A.A. H, Kompa G., Bangert A.* Study about TEM horn size reduction of ultrawideband radar application // 2011 IEEE German Microwave Conference. – 2011. – P. 1-4.
21. *Rojhani N., Pieraccini M., Golazari S.S.* A Compact TEM Horn Antenna for Ground Penetrating Radar // 2018 IEEE International Conference on Advances in Computing, Communications and Informatics (ICACCI). – 2018. – P. 1641-1645. – DOI: 10.1109/ICACCI.2018.8554597.
22. *Savostin V.S., Gevorkyan A.V.* Ultra-Wideband 10-Element Antipodal Vivaldi Antenna Array with Metallic Insert // 2023 Radiation and Scattering of Electromagnetic Waves RSEMW. – 2023. – P. 420-423. – DOI: 10.1109/RSEMW58451.2023.10202131.
23. High Frequency Structural Simulator (HFSS). ANSYS. Available: https://www.ansys.com/products/electronics/ansys-hfss.

REFERENCES

1. *Tayebi M., Dastranj A.A., Alighanbari A.* Ultra Wide Band Antipodal Vivaldi Antenna with Tapered Triangular Corrugated Edges, *IEEE 27th Iranian Conference on Electrical Engineering (ICEE),* 2019, pp. 1637-1642. DOI: 10.1109/IranianCEE.2019.8786722.
2. *Tangwachirapan S., Thaiwirot W., Akkaraekthalin P.* Design of Ultra-Wideband Antipodal Vivaldi Antenna with Square Dielectric Lens for Microwave Imaging Applications, *IEEE 7th International Electrical Engineering Congress (iEECON),* 2019, pp. 1-4. DOI: 10.1109/iEECON45304.2019.8939032.
3. *Gibson P.J.* The Vivaldi Aerial, *IEEE 9th European Microwave Conference,* 1979, pp. 101-105. DOI: 10.1109/EUMA.1979.332681.
4. *Gazit E.* Improved Design of the Vivaldi Antenna. Microwaves, *IEE Proceedings,* 1988, Vol. 135, Pt. H. N. 2, pp. 89-92. DOI: 10.1049/ip-h-2.1988.0020.
5. *Langley J.D.S., Hall P.S., Newham P.* Novel ultrawide-bandwidth Vivaldi antenna and low crosspolarisation, *Electronics Letters,* 1993, Vol. 29, Issue 23, pp. 2004-2005. DOI: 10.1049/el:19931336.
6. *Logan J.T., Kindt R.W., Vouvakis M.N.* A 1.2–12 GHz Sliced Notch Antenna Array, *IEEE Transactions on Antennas and Propagation,* Apr. 2018, Vol. 66, No. 4, pp. 1818-1826. DOI: 10.1109/TAP.2018.2809476.

7. *Ning Y., Ling L., Bao F., Li Z.* An Ultra-wideband Miniaturized Antipodal Vivaldi Antenna, *2018 International Conference on Microwave and Millimeter Wave Technology ICMMT),* 2018, pp. 1-3. DOI: 10.1109/ICACCI.2018.8563335.
8. *Minjie Guo et al.* High-gain antipodal Vivaldi antenna with metamaterial covers, *IET Microwaves, Antennas & Propagation,* 2019, Vol. 13, Iss. 15. DOI: 10.1049/iet-map.2019.0449.
9. *Kumar R., Behera B. R., Suraj P.* A Modified Leaf Shaped Antipodal Vivaldi Antenna for UWB Applications, *2018 IEEE Indian Conference on Antennas and Propogation (InCAP),* 2018, pp. 1-4. DOI: 10.1109/INCAP.2018.8770788.
10. *Suo Y., Qi F., Li W.* Design of ultra-wideband TEM horn antenna for life detection, *2021 IEEE International Symposium on Antenna and Propagation (ISAP),* 2021, pp. 1–2. DOI: 10.23919/ISAP47258.2021.9614515.
11. *Wang D., Zheng S., Deng F., Hou D.* Analysis on the time-domain transfer characteristics of UWB horn antenna, *IEEE 5th International Symposium on Microwave, Antenna, Propagation and EMC Technologies for Wireless Communications,* 2013, pp. 351-354. DOI: 10.1109/MAPE.2013.6689819.
12. *Wang Y., Chen Y.-G.* Research on the impact of TEM horn antenna's structural parameter pulse waveform, *2011 IEEE International Conference on Electrical and Control Engineering,* 2011, pp. 3155-3138. DOI: 10.1109/ICECENG.2011.6057399.
13. *Zalabsky T., Bezousek P.* TEM horn antenna for high energy emission, *IEEE 23rd International Conference Radioelektronika (RADIOELEKTRONIKA),* 2013, pp. 92-95. DOI: 10.1109/RadioElek.2013.6530898.
14. *Lin S., Yu S., Jiao J.-L., Yang C.-T.* Simulation and analysis of an ultra-wideband TEM horn antenna with edge, *2017 IEEE International Symposium on Antennas and Propagation (ISAP),* 2017, pp. 1-2. DOI: 10.1109/ISANP.2017.8228996.
15. *Schoeman K., Meyer P., D.I.L. de Villiers.* Exponential TEM horn with a convex triangular arc, *2013 IEEE Africon,* 2013, pp. 1-5. DOI: 10.1109/AFRCON.2013.6757614.
16. *Kaloshin V.A., Nguen K.Z.* Issledovanie kharakteristik SSHP ploskikh dvumerno-periodicheskikh reshetok TEM ruporov [Study of characteristics of UWB flat two-dimensional periodic arrays of TEM horns], *Zhurnal radioelektroniki* [Journal of Radio Electronics], 2017, No. 5. Available at: http://jre.cplire.ru/jre/may17/14/text.pdf.
17. *Kaloshin V.A., Le N.T., Frolova E.V.* Sverkhdiapazonnaya tsilindricheskaya antennaya reshetka TEM ruporov [Ultra-range cylindrical antenna array of TEM horns], *Zhurnal radioelektroniki* [Journal of Radio Electronics], 2020, No. 4. Available at: https://doi.org/10.30898/1684-1719.2020.4.2.
18. *Suo Y., Qi F., Li W.* Design of exponential gradient TEM horn antenna for ground penetrating radar, *2021 IEEE International Symposium on Antennas and Propagation (ISAP),* 2021, pp. 1-2. DOI: 10.23919/ISAP47258.2021.9614604.
19. *Li Z.-H., Ma J., Wu J., Huang J.-J., Lu Y. -H., Peng L.* Miniaturization Design of TEM Horn Antenna for Ground-Penetrating Radar, *2023 IEEE International Conference on Microwave and Millimeter Wave Technology (ICMMT),* 2023, pp. 1-3. DOI: 10.1109/ICACCI.2018.8554597.
20. *Ameri A.A. H, Kompa G., Bangert A.* Study about TEM horn size reduction of ultrawideband radar application, *2011 IEEE German Microwave Conference,* 2011, pp. 1-4.
21. *Rojhani N., Pieraccini M., Golazari S.S.* A Compact TEM Horn Antenna for Ground Penetrating Radar, *2018 IEEE International Conference on Advances in Computing, Communications and Informatics (ICACCI),* 2018, pp. 1641-1645. DOI: 10.1109/ICACCI.2018.8554597.
22. *Savostin V.S., Gevorkyan A.V.* Ultra-Wideband 10-Element Antipodal Vivaldi Antenna Array with Metallic Insert, *2023 Radiation and Scattering of Electromagnetic Waves RSEMW,* 2023, pp. 420-423. DOI: 10.1109/RSEMW58451.2023.10202131.
23. High Frequency Structural Simulator (HFSS). ANSYS. Available: https://www.ansys.com/products/electronics/ansys-hfss.

Статью рекомендовал к опубликованию д.т.н., профессор К.Е. Румянцев.

**Геворкян Армен Валерьевич** – Южный федеральный университет; e-mail: gevorkyan.scp@yandex.ru; г. Таганрог, Россия; тел.: +78634371733; кафедра АиРПУ; к.т.н.; доцент.

**Савостин Владислав Сергеевич** – e-mail: savostin@sfedu.ru; тел.: +7863437173; кафедра АиРПУ; аспирант.

**Gevorkyan Armen Valerievich** – Southern Federal University; e-mail: gevorkyan.scp@yandex.ru; Taganrog, Russia; phone: +78634371733; the Department of Antennas and Radio Transmitting Devices; cand. of eng. sc.; associate professor.

**Savostin Vladislav Sergeevich** – e-mail: savostin@sfedu.ru; phone: +78634371733; the Department of Antennas and Radio Transmitting Devices; postgraduate student.

УДК 004.052.42 DOI 10.18522/2311-3103-2024-5-271-282

**И.А. Калмыков, И.Д. Ефременков, Д.В. Духовный**

# ПОМЕХОУСТОЙЧИВЫЙ ПРОТОКОЛ ОПОЗНАВАНИЯ НИЗКООРБИТАЛЬНОГО СПУТНИКА-РЕТРАНСЛЯТОРА

*Освоение месторождений в районах Крайнего Севера относится к глобальным проектам, которые реализует Российская Федерация. Эффективный контроль и мониторинг состояния необслуживаемых объектов (НО), занимающихся промыслом углеводородов, достоверное доведение до них команд управления возможно только с помощью низкоорбитальных спутников (НС), объединенных в одну группировку. Однако по мере расширения числа стран, участвующих в разработке месторождений в районах Крайнего Севера, будет расти и количество группировок НС. В результате этого приемник, расположенный на НО, будет видеть сразу несколько спутников-ретрансляторов. При этом НС-злоумышленник (НСЗ) может попытаться навязать приемнику перехваченную ранее команду управления, что может привести к выходу из строя НО. Предотвратить возможность навязывания такой спуфинг-помехи можно с помощью системы опознавания низкоорбитального спутника (СОНС). Эффективность работы СОНС во многом зависит от протокола опознавания. Для повышения скорости проведения аутентификации НС в ряде работ предлагается использовать протокол с нулевым разглашением знаний, который выполняется в модулярных кодах (МК). Данный результат достигается за счет параллельного выполнения арифметических операций по основаниям кода. Однако это свойство МК можно использовать для повышения помехоустойчивости СОНС, которая должна функционировать в различной погодных условиях. Цель – разработка помехоустойчивого протокола опознавания НС-ретранслятора, выполняемого в модулярном коде и требующего меньшего времени на коррекцию ошибок.*

*Протокол опознавания с нулевым разглашением знаний; модулярные коды; поиск и коррекция ошибок; позиционные характеристики.*

**I.A. Kalmykov, I.D. Efremenkov, D.V. Dukhovnyj**

# NOISE-RESISTANT LOW-ORBIT REPEATER SATELLITE IDENTIFICATION PROTOCOL

*The development of deposits in the Far North is one of the global projects implemented by the Russian Federation. Effective control and monitoring of the condition of unattended objects (UO) engaged in hydrocarbon fishing, reliable communication of control commands to them is possible only with the help of low-orbit satellites (LOS) combined into one grouping. However, as the number of countries involved in the development of deposits in the Far North expands, the number of LOS groupings will also grow. As a result, the receiver located on the UO will see several repeater satellites at once. In this case, a low-orbit intruder satellite (LOIS) may try to impose a previously intercepted control command on the receiver, which may lead to the failure of the UO. To prevent the possibility of imposing such spoofing interference, you can use the low-orbit satellite identification system (LSIS). The effectiveness of the LSIS depends largely on the identification protocol. In order to increase the speed of authentication of the NS, in a number of works it is proposed to use a zero-knowledge protocol, which is performed in modular codes (MC). This result is achieved by parallel execution of arithmetic operations on the basis of the code. However, this property of the MC can be used to increase the noise immunity of the LSIS, which must function in various weather conditions. The goal is to develop a noise-resistant protocol for recognizing an LOS-repeater, executed in modular code and requiring less time for error correction.*

*Zero-knowledge recognition protocol; modular codes; error detection and correction; positional characteristics.*

**Введение.** Одним из перспективных глобальных проектов развития Крайнего Севера нашей Родины является разработка и освоение месторождений углеводородов, находящихся в данном регионе. Так как эти месторождения находятся в труднодоступной местности с суровой климатической обстановкой, то добыча и транспортировка углеводородов будет организована с помощью необслуживаемых объектов (НО). Благодаря своим достоинствам, позволяющим повысить эффективность мониторинга, контроля и управления такими необслуживаемыми объектами, снизить вероятности их

выхода из строя, а также уменьшить себестоимость продукции, технологии промышленного интернета вещей (IIoT) нашли широкое применение в нефтегазодобывающую отрасль (НГДО) [1, 2].

Обеспечить достоверное получение информации от НО и доведение до них соответствующих команд управления, возможно только с помощью низкоорбитальных спутников (НС), объединенных в группировку. Перспективность месторождений, находящихся на шельфе Северного Ледовитого океана, стала стимулом к расширению числа стран, участвующих в их разработке. При этом, очевидно, будет расти и количество группировок низкоорбитальных спутников. В результате этого приемники, расположенные на НО, будут видеть в зоне своего приема сразу несколько спутников-ретрансляторов. При этом НС-злоумышленник (НСЗ) может осуществить деструктивное воздействие на необслуживаемый объект. Для этого он сначала осуществляет перехват сигнала (команды), которая передается на НО. Затем он ее, задержав на некоторое время, пытается навязать приемнику. Так как параметры задержанного сигнала являются известными приемнику, то он передает навязанную команду в систему управления НО. В результате этого объект может выйти из строя, и даже вызвать экологическую катастрофу.

Предотвратить возможность навязывания такой спуфинг-помехи можно с помощью системы опознавания низкоорбитального спутника (СОНС). Данная система сначала проводит аутентификацию НС, а затем, если спутник «свой», ему выделяется канал связи. Очевидно, что эффективность работы такой СОНС во многом зависит от применяемого протокола опознавания. В работах [3, 4] представлен протокол опознавания с нулевым разглашением знаний, который обеспечивает аутентификацию НС за меньшее время по сравнению с другими протоколами. Для повышения скорости проведения аутентификации НС в ряде работ [5–7] предлагается выполнять данный протокол в модулярных кодах (МК). Данный результат достигается за счет параллельного выполнения арифметических операций по основаниям кода. При этом независимость остатков МК друг от друга можно использовать для выполнения поиска и коррекции ошибок, возникающих из-за пачек помех. Это свойство МК можно использовать для повышения помехоустойчивости СОНС, которая должна эффективно функционировать в различной погодных условиях. Цель статьи – разработка помехоустойчивого протокола опознавания НС-ретранслятора, выполняемого в модулярном коде и требующего меньшего времени на коррекцию ошибок.

**Постановка задачи исследования.** Кодовая комбинация МК представляет собой кортеж остатков, которые сравнимы с целым числом $X$ по модулю оснований $m_1, m_2, ..., m_k$, где $\underset{i \neq j}{НОД}(m_i, m_j) = 1,\ i,\ j = 1, ..., k$,

$$X = (X_1, X_2, ..., X_k), \tag{1}$$

где $X_i \equiv X \bmod m_i,\ i = 1, ..., k$.

Согласно работам [8–10] в МК эффективно выполняются следующие модульные операции , т.е. сложение, вычитание и умножение

$$X + C = ((X_1 + C_1) \bmod m_1, ..., (X_k + C_k) \bmod m_k), \tag{2}$$

$$X - C = ((X_1 - C_1) \bmod m_1, ..., (X_k - C_k) \bmod m_k), \tag{3}$$

$$X \cdot C = ((X_1 \cdot C_1) \bmod m_1, ..., (X_k \cdot C_k) \bmod m_k), \tag{4}$$

где $C_i = C \bmod m_i;\ i = 1, ..., k$.

Для получения правильного ответа необходимо чтобы результаты выражений (2)–(4) не превышали рабочий диапазон, который равен

$$M_k = \prod_{i=1}^{k} m_i \cdot \tag{5}$$

Анализ выражений (2)–(4) показывает – в МК операции сложения, вычитания и умножения выполняются параллельно, а это способствует повышению скорости вычислений. Это свойство МК было использовано для обеспечения более высокой степени ими-

тостойкости СОНС. Известно [11, 12], что в протоколах аутентификации с нулевым разглашением их имитостойкость определяется не применением шифрования, а за счет выполнения вычислений по модулю большого простого числа *D*. Так как основной операцией в этих протоколах является возведение в степень по модулю *D*, то это негативно сказывается на скорости опознавания НС-ретранслятора. В результате НСЗ появляется дополнительно время на подбор сигнала ответчика, что приводит к снижению имитостойкости СОНС и увеличению вероятности пропуска данного спутника. Чтобы повысить скорость опознавания НС без снижения имитостойкости СОНС в работе [5] были разработаны протоколы аутентификации с нулевым разглашением, который выполнялись в МК. Реализация этих протоколов на ПЛИС Kintex UltraScale (xcku3p-ffva676-1-e) показала, что для выполнения одномодульного протокола Фиат-Шамира потребовалось 21,34 мкс, а при использовании МК – 4,957 мкс. Меньшее время на опознавание, которое составило 3,912 мкс, требует одномодульный бесключевой протокол [3]. Применение МК в данном протоколе позволило снизить временные затраты на опознавание до 1,6 мкс.

Основным недостатком бесключевого протокола в модулярном коде является его низкая помехоустойчивость. Применение МК имело целью за счет распараллеливания вычислений повысить скорость опознавания НС. Однако МК, благодаря тому, что вычисления выполняются независимо по основаниям, могут корректировать ошибочные остатки в кодовых комбинациях. Для этого необходимо ввести в набор оснований дополнительные избыточные модули. Известно [13–15], что расширение набора оснований на два контрольных основания $m_{k+1}, m_{k+2}$ позволяет корректировать однократную ошибку в МК. В этом случае избыточный МК сможет исправлять один ошибочный остаток. Для этого к этим основания предъявляются следующее требование

$$m_{k+1}m_{k+2} > m_{k-1}m_k. \tag{6}$$

Чтобы обеспечить коррекцию пачки ошибок в принятой кодовой комбинации, кратность которой равна $\gamma_{ПО}$, необходимо ввести дополнительно $\rho = 2\gamma_{ПО}$ контрольных оснований. Это приведет к расширению диапазона возможных кодовых комбинаций

$$M_{k+\rho} = \prod_{i=1}^{k+\rho} m_i = M_k \prod_{i=k+1}^{k+\rho} m_i = M_k M_\rho^*. \tag{7}$$

Считается, комбинация МК является разрешенной (она не содержит ошибки), если имеет место неравенство

$$X = (X_1, ..., X_k, X_{k+1}, ..., X_{k+\rho}) < M_k. \tag{8}$$

При наличии в избыточной комбинации МК пачки ошибок кратности $\gamma_{ПО}$ и меньше условие (8) нарушается. Так как основой для классификации принятой комбинации на разрешенные и ошибочные является положение числа *X* относительно рабочего диапазона $M_k$, то для поиска и коррекции ошибок в МК используются позиционные характеристики (ПХ).

Характерной чертой МК является наличие множества алгоритмов вычисления ПХ, с помощью которых можно выполнить поиск и коррекцию ошибок в МК. Одним из первых алгоритмов, позволяющих корректировать ошибки в МК, был разработан алгоритм проекции [16, 17]. Проекция получается из исходной избыточной комбинации $X = (X_1, X_2, X_3, ..., X_{k+\rho})$ путем удаления одного остатка. Так при удалении первого остатка $X_1$ получается проекция $\hat{X}^1 = (X_2, X_3, ..., X_{k+\rho})$. Если удалить второй остаток $X_2$, то получаем вторую проекцию $\hat{X}^2 = (X_1, X_3, ..., X_{k+\rho})$. Процедура выполняется для всех оставшихся остатков. После получения j-ой проекции $\hat{X}^o = (X_1, ..., X_{j-1}, X_{j+1}, ..., X_{k+\rho})$ к ней применяется обратное преобразование из МК в позиционный код (ПК). Для выполнения преобразования МК-ПК часто применяют Китайскую терему об остатках (КТО) [16]

$$\hat{X}^j = \sum_{\substack{i=1 \\ i\neq j}}^{k+\rho} X_i B_i^j \bmod M_{k+\rho}^j = \left| X_1 B_1^j + \ldots + X_{j-1} B_{j-1}^j + X_{j+1} B_{j+1}^j + \ldots + X_{k+\rho} B_{k+\rho}^j \right|_{M_{k+\upsilon}^j}, \quad (9)$$

где $B_i^j = M_{k+\rho}^j w_i \big/ m_i$ – ортогональные базисы;

$M_{k+\rho}^j = M_{k+\rho} \big/ m_j$; $B_i^j = 1 \bmod m_i$; $w_i \left| B_i^j \right|_{m_i}^+ = 1 \bmod m_i$.

Если из-за ошибки исказился *j*-й остаток, то все проекции будут больше рабочего диапазона за исключением текущей проекции

$$\left\{ \hat{X}_1, \ldots, \hat{X}_{j-1}, \hat{X}_{j+1}, \ldots, \hat{X}_{k+\rho} \right\} > M_k, \left\{ \hat{X}_j \right\} < M_k. \quad (10)$$

Основным недостатком данного алгоритма являются большие временные затраты на определение данной позиционной характеристики.

В работах [18, 19] для коррекции ошибок предлагается использовать старшие коэффициенты полиадической системы счисления (ПСС). Известно, что целое число *X* можно представить в виде коэффициентов ПСС $(S_1, S_2, \ldots, S_{k+\rho})$, используя равенство

$$X = S_1 + S_2 m_1 + S_2 m_1 m_2 + \ldots + S_k \prod_{i=1}^{k-1} m_i + S_{k+1} M_k + \ldots + S_{k+\rho} M_k \prod_{j=k+1}^{k+\rho-1} m_j. \quad (11)$$

Анализ выражения (11) показывает, что если выполняется условие (8), т.е. комбинация МК не содержит ошибки, старшие коэффициенты ПСС должны иметь значение $S_{k+1} = 0, S_{k+2} = 0, \ldots, S_{k+\rho} = 0$. Если эти коэффициенты будут отличаться от нуля, то это означает, что комбинация МК является ошибочной. Для вычисления коэффициентов ПСС, используя модулярный код, в работе [18] был предложен итерационный алгоритм

$$\begin{aligned} S_1 &= X_1, \\ S_2 &= ((X_2 - S_1) u_{12}) \bmod m_2, \\ &\vdots \\ S_{k+\rho} &= ((S_{k+\rho} + S_1) u_{1k+\rho} + S_2) u_{2k+\rho} + \ldots + S_{k+\rho=1} u_{(k+\rho+1)(k+\rho)}) \bmod m_{k+\rho}, \end{aligned} \quad (12)$$

где $u_{ji} = \left(1 \big/ m_j\right) \bmod m_i = m_j^{-1} \bmod m_i$; $i = 1, \ldots, k+\rho$.

В качестве недостатка алгоритма (12) перевода МК-ПСС можно выделить значительные временные затраты из-за итерационного процесса выполнения.

Довольно часто в алгоритмах коррекции ошибок в МК применяется ПХ – интервал числа [19, 20]. Данная ПХ имеет простой физический смысл, определяемый выражением

$$L = \left[ X / M_k \right], \quad (13)$$

где [*] – целая часть частного.

Если комбинация МК является разрешенной, то выполняется условие (8). Так как число *X* меньше рабочего диапазона, то *L* = 0. Если комбинация МК содержит ошибки, то по интервалу $L \neq 0$ можно провести их коррекцию. В работе [21] авторы предлагают вычислять *L* с помощью функции Эйлера

$$L = \left| \sum_{i=1}^{k+\rho} \left| S_{X_i} X_i \right|_{M_k}^+ \right|_{M_k}^+, \quad (14)$$

где $S_{X_i} = \left| M_i^{\varphi(m_i)} \left( M_{k+\rho} \right)^{-1} \right|_{M_k}^+$; $\varphi(m_i)$ – функция Эйлера числа $m_i$; $M_i = \dfrac{M_{k+\rho}}{m_i}$.

К основным недостаткам текущего алгоритма можно отнести большие временные затраты на вычисление ПХ. Уменьшить время необходимое на нахождение интервала числа позволяет разработанный алгоритм коррекции пачек ошибок. В основе алгоритма лежит изоморфизм КТО. Если в выражение (13) подставить (9), используемое для выполнения преобразования МК-ПК, то

$$L = \left[ \sum_{i=1}^{k+\rho} X_i B_i - r_X M_{k+\rho} \Big/ M_k \right], \tag{15}$$

где $r_x$ – ранг числа *X,* представленного по информационным основаниям МК.

Представим ортогональные базисы оснований $m_1, m_2, ..., m_k$ в виде

$$B_i = \left[ B_i / M_k \right] \cdot M_k + \ddot{B}_i = K_i M_k + \ddot{B}_i, \tag{16}$$

где $\ddot{B}_i$ – ортогональные базисы МК с основаниями $m_1, ..., m_k$; $\ddot{B}_i \equiv B_i \bmod M_k$; $i = 1,2,...,k$.

Для контрольных оснований, где $i = k+1,...,k+\rho$, справедливо равенство

$$B_i = K_i M_k. \tag{17}$$

Для перехода от вычисления целой части частого в (15) к модульным операциям воспользуемся свойством полного диапазона избыточного МК. Из выражения (7) видно, что полный диапазон $M_{k+\rho}$ больше рабочего диапазона разрешенных комбинаций $M_k$ в $M_\rho^*$ раз. Другими словами в состав $M_{k+\rho}$ входит $M_\rho^*$ рабочих диапазонов $M_k$. При этом каждый рабочий диапазон имеет свой номер от 0 до $M_\rho^* - 1$. Значит равенство (15) можно представить как

$$L = \left| \sum_{i=1}^{k+\rho} X_i B_i - r_X M_{k+\rho} \Big/ M_k \right|_{M_\rho^*}^{+}. \tag{18}$$

Подставим выражения (16) и (17) в последнее выражение. Тогда

$$L = \left| \sum_{i=1}^{k+\rho} X_i K_i + \left[ \frac{\sum_{j=1}^{k} X_j \ddot{B}_j}{M_k} \right] \right|_{M_\rho^*}^{+} = \left| \sum_{i=1}^{k+\rho} X_i K_i + \ddot{r}_X \right|_{M_\rho^*}^{+}, \tag{19}$$

где $\ddot{r}_X$ – ранг числа *X*, представленного в МК с основаниями $m_1, m_2, ..., m_k$.

Алгоритм, описываемый выражением (19) имеет ряд недостатков. Во-первых, вычисление ПХ осуществляется по модулю, который представляет собой произведение контрольных оснований $M_\rho^* = \prod_{ш=k+1}^{k+\rho} m_i$. Это влечет за собой увеличение аппаратурных затрат. Во-вторых, чтобы вычислить ПХ необходимо выполнить $k+\rho$ операций умножения остатков кода на константы $K_i$, где $i = 1,2,...,k+\rho$ и столько же операций сложения по модулю $M_\rho^*$.

Для устранения отмеченных было предложено использовать изоморфизм Китайской теоремы об остатках. Данное свойство КТО позволяет свети вычисление выражения (19) к $\rho$ параллельным вычислений ПХ по контрольным модулям. Получаем

$$L_{k+1} = |L|^{+}_{m_{k+1}} = \left| \sum_{i=1}^{k} X_i \left| K_i \right|^{+}_{m_{k+1}} + X_{k+1} \left| K_{k+1} \right|_{m_{k+1}} + \ddot{\ddot{r}}_X \right|^{+}_{m_{k+1}},$$
$$\vdots \qquad (20)$$
$$L_{k+\rho} = |L|^{+}_{m_{k+\rho}} = \left| \sum_{i=1}^{k} X_i \left| K_i \right|^{+}_{m_{k+\rho}} + X_{k+\rho} \left| K_{k+\rho} \right|_{m_{k+\rho}} + \ddot{\ddot{r}}_X \right|^{+}_{m_{k+\rho}}.$$

Следует отметить, что на передающей стороне СОНС для построения помехоустойчивого кода необходимо реализовать процесс вычисления контрольных остатков в МК. Для этого был разработан алгоритм вычисления контрольных остатков. Положим, что имеется набор оснований $m_1,...,m_k,m_{k+1}$. Если комбинация не содержит ошибки, то справедливо равенство

$$L_{k+1} = \left| \sum_{i=1}^{k} X_1 \left| K_1 \right|_{m_{k+1}} + X_{k+1} \left| K_{k+11} \right|_{m_{k+1}} + \ddot{\ddot{r}}_X \right|^{+}_{m_{k+1}} = 0. \qquad (21)$$

Выполнив преобразования, получаем алгоритм вычисления первого контрольного остатка

$$X_{k+1} = \left| m_{k+1} - \left( K_{m+1} \right)^{-1} \left| \sum_{i=1}^{k} X_i \left| K_i \right|_{m_{k+1}} + \ddot{\ddot{r}}_X \right|^{+}_{m_{k+1}} \right|^{+}_{m_{k+1}}. \qquad (22)$$

Остальные контрольные можно получить аналогично.

Алгоритмы вычисления контрольных оснований (22) и коррекции ошибок в МК (20) были использованы в разработанном помехоустойчивом протоколе опознавания с нулевым разглашением знаний, вычисления в котором выполняются в МК. Данный протокол состоит из нескольких частей.

**Предварительная часть помехоустойчивого протокола:**

1. Генерируется набор, состоящий из *k* оснований $m_1$, $m_2,...,m_k$ МК, для которого справедливо неравенство $M_k > D$, где D – большое простое число, по модулю которого выполняется протокол опознавания [3].

2. Представление секретных элементов протокола в модулярном коде: $U = (U_1, U_2,...,U_k)$ – секретный ключ спутника; $S(j) = (S_1(j),...,S_k(j))$ – сеансовый ключ; $T(j) = (T_1(j),...,T_k(j))$ – параметр, который необходим для проверки повторного использования *S*(*j*), где *j* – номер сеанса.

3. В зависимости от кратности пачек ошибок $\gamma_{ПО}$ происходит выбор контрольных оснований $m_{k+1},...,m_{k+\rho}$.

**Основная часть помехоустойчивого протокола**. Первый этап основной части протокола включает себя процедуры:

1. Ответчик СОНС, который располагается на спутнике, получает истинный статус НС, представленный в МК

$$C_i = \left| g^{U_1} g^{S_1(j)} g^{T_1(j)} \right|^{+}_{m_i}, \qquad (23)$$

где *g* – порождающий элемент по модулю $m_i$; $i = 1,2,\ldots, k$.

Используя алгоритм (22), вычисляет контрольные остатки для истинного статуса НС. Комбинация $(C_1,...,C_k,C_{k+1},...,C_{k+\rho})$ хранится в блоке памяти НС.

2. Ответчик выбирает случайные комбинации $\Delta U(j)=(\Delta U_1(j),..\Delta U_k(j))$, $\Delta S(j)=(\Delta S_1(j),..\Delta S_k(j))$, $\Delta T(j)=(\Delta T_1(j),..\Delta T_k(j))$, с помощью которых проводит «зашумление» секретных элементов помехоустойчивого протокола

$$U_i^*(j)=\left|U_i+\Delta U_i(j)\right|_{\varphi(m_i)}^{+};\ S_i^*(j)=\left|S_i(j)+\Delta S_i(j)\right|_{\varphi(m_i)}^{+};$$

$$T_i^*(j)=\left|T_i(j)+\Delta T_i(j)\right|_{\varphi(m_i)}^{+}. \quad (24)$$

3. Ответчик получает «зашумленный» статус НС в МК

$$C_i^*=\left|g^{U_i^*(j)}g^{S_i^*(j)}g^{T_i^*(j)}\right|_{m_i}^{+}. \quad (25)$$

Используя алгоритм (22), вычисляет контрольные остатки для зашумленного статуса НС. Комбинация $(C_1^*,...,C_k^*,C_{k+1}^*,...,C_{k+\rho}^*)$ хранится в блоке памяти НС.

На втором этапе протокола опознавания выполняются процедуры.

1. Запросчик СОНС, находящийся на НО, генерирует случайное число *d*. Это число представляет в виде $d=(d_1,...,d_{k+\rho})$, а затем передается ответчику.

2. Ответчик, используя алгоритм (20), сначала проверяет на наличие ошибок полученный «вопрос» $d=(d_1,d_2,...,d_{k+\rho})$. А затем готовит три ответа

$$r_i^1(j)=\left|U_i^*(j)-d_iU_i\right|_{\varphi(m_i)}^{+};\ r_i^2(j)=\left|S_i^*(j)-d_iS_i(j)\right|_{\varphi(m_i)}^{+};$$

$$r_i^3(j)=\left|T_i^*(j)-d_iT_i(j)\right|_{\varphi(m_i)}^{+}. \quad (26)$$

Используя алгоритм (22), ответчик вычисляет контрольные остатки для трех ответов. Затем формируется сигнал ответчика, который передается запросчику

$$\left\{(C_1,...,C_{k+\rho}),(C_1^*,...,C_{k+\rho}^*),(r_1^1(j),...,r_{k+\rho}^1(j)),(r_1^2(j),...,r_{k+\rho}^2(j)),(r_1^3(j),...,r_{k+\rho}^3(j))\right\}.$$

3. Запросчик с помощью алгоритма (20) осуществляет поиск и коррекцию пачек ошибок в принятом сигнале. Затем он приступает к проверке сигнала

$$Y_i=\left|(C_i)^{d_i}g^{r_i^1(j)}g^{r_i^2(j)}g^{r3_i(j)}\right|_{m_i}^{+}. \quad (27)$$

При выполнении условия

$$\{Y_1=C_1^*,Y_2=C_2^*,...,Y_k=C_k^*\}, \quad (28)$$

спутник получает статус «свой», и СОНС предоставляет ему сеанс связи.

**Результаты экспериментальных исследований.** Рассмотрим пример работы разработанного помехоустойчивого протокола, осуществляющего опознавание с использованием МК.

**Предварительная часть помехоустойчивого протокола:**

1. Пусть имеем простое число *D* = 2251. Выбираем информационные основания $m_1=7, m_2=17, m_3=19$, так как их рабочий диапазон $M_3=2261$ превышает число *D*. При этом они имеют одинаковый элемент *g* = 2.

2. Представим секретные элементы протокола в модулярном коде: секретный ключ НС $U=1000=(3,14,12)$; сеансовый ключ $S(j)=248=(3,10,1)$; параметр проверки $T(j)=452=(4,10,15)$.

3. Для более простых расчетов выберем кратность исправляемой ошибки $\gamma=1$. Тогда достаточно добавить два контрольных основания $m_4=29, m_5=37$.

**Основная часть помехоустойчивого протокола**. Первый этап основной части протокола включает себя процедуры:

1. Получение ответчиком СОНС истинного статуса НС согласно (23)

$$C_1 = \left| g^{U_1} g^{S_1(j)} g^{T_1(j)} \right|_{m_1}^{+} = \left| 2^3 \cdot 2^3 \cdot 2^4 \right|_7^{+} = \left| 2^4 \right|_7^{+} = 2.$$

$$C_2 = \left| g^{U_2} g^{S_2(j)} g^{T_2(j)} \right|_{m_2}^{+} = \left| 2^{14} \cdot 2^{10} \cdot 2^{10} \right|_{17}^{+} = \left| 2^2 \right|_{17}^{+} = 4.$$

$$C_3 = \left| g^{U_3} g^{S_3(j)} g^{T_3(j)} \right|_{m_3}^{+} = \left| 2^{12} \cdot 2^1 \cdot 2^{15} \right|_{19}^{+} = \left| 2^{10} \right|_{19}^{+} = 17.$$

Используя алгоритм (22), вычислим первый контрольный остаток. Для полной системы оснований вычислим ортогональные базисы

$$B_1 = 1386316 = K_1 \cdot M_3 + \ddot{B}_1 = 613 \cdot 2261 + 323 \cdot$$

$$B_2 = 1997926 = K_2 \cdot M_3 + \ddot{B}_2 = 883 \cdot 2261 + 1463 \cdot$$

$$B_3 = 1404557 = K_3 \cdot M_3 + \ddot{B}_3 = 621 \cdot 2261 + 476 \cdot$$

$$B_4 = 1505826 = K_4 \cdot M_3 = 666 \cdot 2261 \cdot$$

$$B_5 = 983535 = K_5 \cdot M_3 = 435 \cdot 2261 \cdot$$

Имеем безизбыточную комбинацию $C(j) = (2, 4, 17)$. Вычислим ранг числа

$$\ddot{r}_X = \left\lfloor \frac{\sum_{i=1}^{3} X_i \ddot{B}_i}{M_3} \right\rfloor = \left\lfloor \frac{2 \cdot 323 + 4 \cdot 1463 + 17 \cdot 476}{2261} \right\rfloor = 12 \cdot$$

Тогда первый контрольный остаток равен

$$X_4 = \left| m_4 - \frac{\left| \sum_{i=1}^{k} X_i \left| K_i \right|_{m_{k+1}} + \ddot{r}_U \right|_{m_4}}{K_4} \right|_{m_4}^{+} = \left| 29 - \frac{2 \cdot 4 + 4 \cdot 13 + 17 \cdot 12}{28} + 12 \right|_{29}^{+} = 0 \cdot$$

Аналогичным образом получили $X_5 = 16$. Тогда $C(j) = (2,4,17,0,16) \cdot$

2. Ответчик выбирает $\Delta U(j) = (5,13,5)$, $\Delta S(j) = (5,5,11)$, $\Delta T(j) = (1,13,7)$, с помощью которых проводит «зашумление» секретных элементов

$$U^*(j) = \left( \left| 3+5 \right|_6^{+}, \left| 14+13 \right|_{16}^{+}, \left| 12+5 \right|_{18}^{+} \right) = (2,11,17); \; S^*(j) = (2,15,12); \; T^*(j) = (5,7,4) \cdot$$

3. Ответчик получает «зашумленный» статус НС в МК согласно (25)

$$C_1^* = \left| g^{U_1^*(j)} g^{S_1^*(j)} g^{T_1^*(j)} \right|_{m_1}^{+} = \left| 2^2 \cdot 2^2 \cdot 2^5 \right|_7^{+} = \left| 2^3 \right|_7^{+} = 1.$$

$$C_2^* = \left| g^{U_2^*(j)} g^{S_2^*(j)} g^{T_2^*(j)} \right|_{m_2}^{+} = \left| 2^{11} \cdot 2^{15} \cdot 2^7 \right|_{17}^{+} = \left| 2^1 \right|_{17}^{+} = 2.$$

$$C_3^* = \left| g^{U_3^*(j)} g^{S_3^*(j)} g^{T_3^*(j)} \right|_{m_3}^{+} = \left| 2^{17} \cdot 2^{12} \cdot 2^4 \right|_{19}^{+} = \left| 2^{15} \right|_{19}^{+} = 12.$$

Используя алгоритм (22), получили $C^*(j) = (1,2,12,3,32) \cdot$

На втором этапе протокола опознавания выполняются процедуры.

1. Запросчик СОНС, находящийся на НО, генерирует случайное число $d = (3,16,6,14,27)$, которое затем передал ответчику

2. Ответчик сначала проверяет на наличие ошибок полученный «вопрос». Пусть при передаче помеха исказила первый остаток, и принятая ошибочная комбинация имеет вид $\widetilde{d} = (\widetilde{0},16,6,14,27)$. Воспользуемся алгоритмом (20). Вычислим ранг числа

$$\ddot{r}_d = \left[\frac{\sum_{i=1}^{3} d_i \ddot{B}_i}{M_3}\right] = \left[\frac{0\cdot 323 + 16\cdot 1463 + 6\cdot 476}{2261}\right] = 11.$$

Вычислим интервал числа *d*, представленного в модулярном коде

$$L_4 = \left|\sum_{i=1}^{3} d_i \left|K_i\right|_{m_4} + d_4\left|K_4\right|_{m_4} + \ddot{r}_d\right|_m^+ = \left|0\cdot 4 + 16\cdot 13 + 6\cdot 12 + 14\cdot 28 + 11\right|_{29}^+ = 16.$$

$$L_5 = \left|\sum_{i=1}^{3} d_i \left|K_i\right|_{m_5} + d_5\left|K_5\right|_{m_5} + \ddot{r}_d\right|_{m_5}^+ = \left|0\cdot 21 + 16\cdot 32 + 6\cdot 29 + 27\cdot 28 + 11\right|_{29}^+ = 10.$$

Так позиционная характеристика не равна нулю, то это означает, что комбинация содержит ошибку. Для значений $L_4 = 16, L_5 = 10$ вектор ошибки равен $\bar{e} = (3,0,0,0,0)$. Произведем исправление ошибки

$$d = \widetilde{d} + \bar{e} = \widetilde{d} = (\widetilde{0},16,6,14,27) + (3,0,0,0,0) = \widetilde{d} = (3,16,6,14,27).$$

Затем ответчик готовит три ответа согласно (26)

$$r_1^1(j) = \left|U_1^*(j) - d_1 U_1\right|_{\varphi(m_1)}^+ = \left|2 - 3\cdot 3\right|_6^+ = 5.\ r_1^2(j) = \left|2 - 3\cdot 3\right|_6^+ = 5.\ r_1^3(j) = \left|5 - 3\cdot 4\right|_6^+ = 5.$$

$$r_2^1(j) = \left|U_2^*(j) - d_2 U_2\right|_{\varphi(m_2)}^+ = \left|11 - 16\cdot 14\right|_{16}^+ = 11.\ r_2^2(j) = \left|15 - 16\cdot 10\right|_{16}^+ = 15.$$

$$r_2^3(j) = \left|7 - 16\cdot 10\right|_{16}^+ = 7.$$

$$r_3^1(j) = \left|U_3^*(j) - d_3 U_3\right|_{\varphi(m_3)}^+ = \left|17 - 6\cdot 12\right|_{18}^+ = 17.\ r_3^2(j) = \left|12 - 6\cdot 1\right|_{18}^+ = 6.$$

$$r_3^3(j) = \left|4 - 6\cdot 15\right|_{18}^+ = 4.$$

Используя алгоритм (22), ответчик вычисляет контрольные остатки для трех ответов. Затем формируется сигнал ответчика, который содержит: истинный статус $C(j) = (2,4,17,0,16)$, «зашумленный статус» $C^*(j) = (1,2,12,3,32)$, ответа $r^1(j) = (5,11,17,1,4)$, $r^2(j) = (5,15,6,8,28)$, $r^3(j) = (5,7,4,20,9)$. Сигнал передается запросчику.

3. Запросчик с помощью алгоритма (20) проверил принятый сигнал. Пусть в нем ошибок не было. Затем он приступает к проверке сигнала

$$Y_1 = \left|\left(C_1\right)^{d_1} g^{r_1^1(j)} g^{r_1^2(j)} g^{r_1^3(j)}\right|_{m_1}^+ = \left|2^3\cdot 2^5\cdot 2^5\cdot 2^5\right|_7^+ = \left|2^0\right|_7^+ = 1.$$

$$Y_2 = \left|\left(C_2\right)^{d_2} g^{r_2^1(j)} g^{r_2^2(j)} g^{r_2^3(j)}\right|_{m_2}^+ = \left|4^{16}\cdot 2^{11}\cdot 2^{15}\cdot 2^7\right|_{17}^+ = \left|2^1\right|_{17}^+ = 2.$$

$$Y_3 = \left|\left(C_3\right)^{d_3} g^{r_3^1(j)} g^{r_3^2(j)} g^{r_3^3(j)}\right|_{m_3}^+ = \left|17^6\cdot 2^{17}\cdot 2^6\cdot 2^4\right|_{17}^+ = \left|2^{15}\right|_{19}^+ = 12.$$

Так как условие(28) выполнилось

$$\left\{Y_1 = C_1^* = 1, Y_2 = C = 2_2^*, Y_3 = C_k^* = 12\right\}, \tag{28}$$

спутник получил статус «свой», и СОНС предоставила ему сеанс связи.

В ходе исследований был проведен анализ двух алгоритмов коррекции ошибок в МК с использованием FPGA Xilinx Artix-7 (xc7a12ticsg325-1L). Результаты моделирования показали, что при выполнении алгоритма (19) на поиск и исправление ошибок было потрачено 152 нс. Применение разработанного алгоритма (20) сократило это время до 123 нс. Поставленная цель достигнута.

**Выводы.** В статье представлен разработанный помехоустойчивый протокол опознавания низкоорбитальных спутников., который выполняется в МК. Применение модулярного кода позволяет не только уменьшить временные затраты на опознавание НС, но и способствует приданию СОНС свойства помехоустойчивости. Проведен анализ алгоритмов поиска и коррекции ошибок в МК. Показано, что данные алгоритмы имеют большие временные затраты. Для устранения этого недостатка был разработан алгоритм коррекции ошибок в МК, использующий изоморфизм КТО. Также этот алгоритм был использован для процедуры вычисления контрольных остатков, которая реализуется перед сигналов по каналу связи. Данные алгоритмы были использованы в разработанном помехоустойчивом протоколе опознавания низкоорбитального спутника-ретранслятора. В ходе исследований был проведен анализ двух алгоритмов коррекции ошибок в МК с использованием FPGA Xilinx Artix-7 (xc7a12ticsg325-1L). Результаты моделирования показали, что при выполнении алгоритма (19) на поиск и исправление ошибок было потрачено 152 нс. Применение разработанного алгоритма (20) сократило это время до 123 нс. Таким образом, разработанный алгоритм коррекции оказывает меньшее влияние на скорость опознавания НС. В результате этого СОНС, использующая разработанный помехоустойчивый алгоритм будет обладать более высокой имитостойкостью по сравнению с алгоритмом (19).

*Исследование выполнено за счет гранта Российского научного фонда, grant number 23-21-00036, https://rscf.ru/en/project/23-21-00036/.*

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Прахова М.Ю., Хорошавина Е.А.* Системы автоматизации в нефтяной промышленности. – М. – Вологда: Инфра-Инженерия, 2019. – 304 с.
2. *Liao Y., Loures E. de F.R., Deschamps F.* Industrial internet of things: a systematic literature review and insights // IEEE Internet Things Journal. – 2018. – Vol. 5 (6). – P. 4515-4525. – DOI: 10.1109/JIOT.2018.2834151.
3. *Kalmykov I.A., Olenev A.A., Kalmykova N.I., Dukhovnyj D.V.* Using Adaptive Zero-Knowledge Authentication Protocol in VANET Automotive Network // Information. – 2023. – Vol. 14(1), 27. – DOI: 10.3390/info14010027.
4. *Chistousov N.K., Kalmykov I.A., Dukhovnyj D.V., Kalmykov M.I., Olenev A.A.* Adaptive Authentication Protocol Based on Zero-Knowledge Proof // Algorithms. – 2022. – Vol. 15 (2), 50. – DOI: 10.3390/a15020050.
5. *Olenev A.A., Kalmykov I.A., Pashintsev V.P.* Improved Spacecraft Authentication Method for Satellite Internet System Using Residue Codes // Information. – 2023. – Vol. 14 (7), 407. – DOI: 10.3390/info14070407.
6. *Kalmylov I.A., Kopytov V.V., Olenev A.A., Kalmykova N.I., Chistousov N.K.* Application of Modular Residue Classes Codes in an Authentication Protocol for Satellite Internet Systems // IEEE Access. – 2023. – Vol. 11:1-1. – P. 71624-71633. – DOI: 10.1109/ACCESS.2023.3290498.
7. *Чистоусов Н.К., Калмыков И.А., Чипига А.Ф., Калмыкова Н.И.* Разработка протоколов аутентификации низкоорбитальных космических аппаратов на основе параллельных кодов систем остаточных классов // Инженерный вестник Дона. – 2021. – № 4. – URL: ivdon.ru/ru/magazine/archive/n4y2021/6912.
8. *Tyncherov K.T., Mukhametshin V.Sh., Khuzina L.B.* Method to control and correct telemetry well information in the basis of residue number system // Journal of Fundamental and Applied Sciences. – 2017. – Vol. 9 (2). – P. 1370-1374.
9. *Mohan A*. Residue Number Systems. Theory and Applications. – Switzerland, Springer International Publishing, 2016. – 351 p.
10. *Mohan A*. RNS-Based arithmetic circuits and applications // Arithmetic Circuits for DSP Applications. Chapter 6. / Eds. P.K. Meher, T. Stouraitis. – John Wiley and Sons, Ltd, 2017. – P. 186-236.
11. *Черемушкин А.В.* Криптографические протоколы. Основные свойства и уязвимости. – М.: Академия, 2009. – 272 с.

12. *Запечников С.В.* Криптографические протоколы и их применение в финансовой и коммерческой деятельности. – М.: Горячая линия-Телеком, 2011. – 256 с.
13. *Omondi A., Premkumar B.* Residue Number Systems: Theory and Implementation. – United Kingdom, Imperial College Press, 2007. – 293 p.
14. *Червяков Н.И., Коляда А.А., Ляхов П.А.* Модулярная арифметика и ее приложения в инфокоммуникационных технологиях. – М.: Физматлит, 2017. – 400 с.
15. *Червяков Н.И., Нагорнов Н.Н.* Коррекция ошибок при передаче и обработке информации, представленной в СОК, методом синдромного декодирования // Наука. Инновации. Технологии. – 2015. – № 2. – С. 15-40.
16. *Акушский И.Я., Юдицкий Д.М.* Машинная арифметика в остаточных классах. – М.: Сов. радио, 1968. – 440 с.
17. *Сиора А.А., Краснобаев В.А., Харченко В.С.* Отказоустойчивые системы с версионно-информационной избыточностью. – Харьков: ХАИ, 2009. – 321 с.
18. *Sun J.-D., Krishna H.* A coding theory approach to error control in redundant residue number systems. II. Multiple error detection and correction // IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing. – 1992. – Vol. 39 (1). – P. 18-34.
19. *Червяков Н.И., Сахнюк П.А., Макоха А.Н.* Нейрокомпьютеры в остаточных классах. Кн. 11. – М.: Радиотехника, 2003. – 272 с.
20. *Chervyakov N.I., Lyakhov P.A., Babenko M.G., Lavrinenko I.N., Lavrinenko A.V., Nazarov A.S.* The architecture of a fault-tolerant modular neurocomputer based on modular number projections // Neurocomputing. – 2018. – Vol. 10. – P. 96-107.
21. *Червяков Н.И., Сахнюк П.А., Шапошников А.В., Ряднов С.А.* Модулярные параллельные вычислительные структуры нейропроцессорных систем. – М.: Физматлит, 2003. – 288 с.

REFERENCES

1. *Prakhova M.Yu., Khoroshavina E.A.* Sistemy avtomatizatsii v neftyanoy promyshlennosti [Automation systems in the oil industry]. Moscow – Vologda: Infra-Inzheneriya, 2019, 304 p.
2. *Liao Y., Loures E. de F.R., Deschamps F.* Industrial internet of things: a systematic literature review and insights, *IEEE Internet Things Journal,* 2018, Vol. 5 (6), pp. 4515-4525. DOI: 10.1109/JIOT.2018.2834151.
3. *Kalmykov I.A., Olenev A.A., Kalmykova N.I., Dukhovnyj D.V.* Using Adaptive Zero-Knowledge Authentication Protocol in VANET Automotive Network, *Information,* 2023, Vol. 14(1), 27. DOI: 10.3390/info14010027.
4. *Chistousov N.K., Kalmykov I.A., Dukhovnyj D.V., Kalmykov M.I., Olenev A.A.* Adaptive Authentication Protocol Based on Zero-Knowledge Proof, *Algorithms,* 2022, Vol. 15 (2), 50. DOI: 10.3390/a15020050.
5. *Olenev A.A., Kalmykov I.A., Pashintsev V.P.* Improved Spacecraft Authentication Method for Satellite Internet System Using Residue Codes, *Information,* 2023, Vol. 14 (7), 407. DOI: 10.3390/info14070407.
6. *Kalmylov I.A., Kopytov V.V., Olenev A.A., Kalmykova N.I., Chistousov N.K.* Application of Modular Residue Classes Codes in an Authentication Protocol for Satellite Internet Systems, *IEEE Access,* 2023, Vol. 11:1-1, pp. 71624-71633. DOI: 10.1109/ACCESS.2023.3290498.
7. *Chistousov N.K., Kalmykov I.A., Chipiga A.F., Kalmykova N.I.* Razrabotka protokolov autentifikatsii nizkoorbital'nykh kosmicheskikh apparatov na osnove parallel'nykh kodov sistem ostatochnykh klassov [Development of authentication protocols for low-orbit spacecraft based on parallel codes of residual class systems], *Inzhenernyy vestnik Dona* [Engineering Bulletin of the Don], 2021, No. 4. Available at: ivdon.ru/ru/magazine/archive/n4y2021/6912.
8. *Tyncherov K.T., Mukhametshin V.Sh., Khuzina L.B.* Method to control and correct telemetry well information in the basis of residue number system, *Journal of Fundamental and Applied Sciences,* 2017, Vol. 9 (2), pp. 1370-1374.
9. *Mohan A*. Residue Number Systems. Theory and Applications. Switzerland, Springer International Publishing, 2016, 351 p.
10. *Mohan A*. RNS-Based arithmetic circuits and applications, *Arithmetic Circuits for DSP Applications. Chapter 6*, Eds. P.K. Meher, T. Stouraitis. John Wiley and Sons, Ltd, 2017, pp. 186-236.
11. *Cheremushkin A.V.* Kriptograficheskie protokoly. Osnovnye svoystva i uyazvimosti [Cryptographic protocols. Basic properties and vulnerabilities]. Moscow: Akademiya, 2009, 272 p.
12. *Zapechnikov S.V.* Kriptograficheskie protokoly i ikh primenenie v finansovoy i kommercheskoy deyatel'nosti [Cryptographic protocols and their application in financial and commercial activities]. Moscow: Goryachaya liniya-Telekom, 2011, 256 p.

13. *Omondi A., Premkumar B.* Residue Number Systems: Theory and Implementation. United Kingdom, Imperial College Press, 2007, 293 p.
14. *Chervyakov N.I., Kolyada A.A., Lyakhov P.A.* Modulyarnaya arifmetika i ee prilozheniya v infokommunikatsionnykh tekhnologiyakh [Modular arithmetic and its applications in infocommunication technologies]. Moscow: Fizmatlit, 2017, 400 p.
15. Chervyakov N.I., Nagornov N.N. Korrektsiya oshibok pri peredache i obrabotke informatsii, predstavlennoy v SOK, metodom sindromnogo dekodirovaniya [Error correction in transmission and processing of information presented in the SOC using the syndrome decoding method], *Nauka. Innovatsii. Tekhnologii* [Science. Innovations. Technologies], 2015, No. 2, pp. 15-40.
16. *Akushskiy I.Ya., Yuditskiy D.M.* Mashinnaya arifmetika v ostatochnykh klassakh [Machine arithmetic in residual classes]. Moscow: Sov. radio, 1968, 440 p.
17. *Siora A.A., Krasnobaev V.A., Kharchenko V.S.* Otkazoustoychivye sistemy s versionno-informatsionnoy izbytochnost'yu [Fault-tolerant systems with version-information redundancy]. Khar'kov: KhAI, 2009, 321 p.
18. *Sun J.-D., Krishna H.* A coding theory approach to error control in redundant residue number systems. II. Multiple error detection and correction, *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing,* 1992, Vol. 39 (1), pp. 18-34.
19. *Chervyakov N.I., Sakhnyuk P.A., Makokha A.N.* Neyrokomp'yutery v ostatochnykh klassakh [Neurocomputers in residual classes]. Book 11. Moscow: Radiotekhnika, 2003, 272 p.
20. *Chervyakov N.I., Lyakhov P.A., Babenko M.G., Lavrinenko I.N., Lavrinenko A.V., Nazarov A.S.* The architecture of a fault-tolerant modular neurocomputer based on modular number projections, *Neurocomputing,* 2018, Vol. 10, pp. 96-107.
21. *Chervyakov N.I., Sakhnyuk P.A., Shaposhnikov A.V., Ryadnov S.A.* Modulyarnye parallel'nye vychislitel'nye struktury neyroprotsessornykh system [Modular parallel computing structures of neuroprocessor systems]. Moscow: Fizmatlit, 2003, 288 p.

Статью рекомендовала к опубликованию д.т.н., профессор Л.К. Бабенко.

**Калмыков Игорь Анатольевич** – Северо-Кавказский федеральный университет; e-mail: kia762@yandex.ru; г. Ставрополь, Россия; тел.: 89034163533; кафедра вычислительной математики и кибернетики; д.т.н.; профессор.

**Ефременков Иван Дмитриевич** – e-mail: kia545@yandex.ru; тел.: 89283097648; кафедра вычислительной математики и кибернетики; старший преподаватель.

**Духовный Даниил Вячеславович** – e-mail: dduhovny26@gmail.com; тел.: 89620070023; кафедра вычислительной математики и кибернетики; аспирант.

**Kalmykov Igor Anatolyevich** – North Caucasus Federal University; e-mail: kia762@yandex.ru; Stavropol, Russia; phone: +79034163533; the Department of Computational Mathematics and Cybernetics; dr. of eng. sc.; professor.

**Efremenkov Ivan Dmitrievich** – e-mail: kia545@yandex.ru; phone: +79283097648; the Department of Computational Mathematics and Cybernetics; senior lecturer.

**Dukhovnyj Daniil Vyacheslavovich** – e-mail: dduhovny26@gmail.com; phone: +79620070023; the Department of Computational Mathematics and Cybernetics; postgraduate student.

# Раздел IV. Сообщение об отзыве публикации

**Глод Ольга Денисовна**

**МОДЕЛЬ СЕТЕВОГО ПЛАНИРОВАНИЯ И УПРАВЛЕНИЯ В НЕЧЕТКИХ ИНТЕРВАЛЬНЫХ ОЦЕНКАХ**

*Глод Ольга Денисовна.* Модель сетевого планирования и управления в нечетких интервальных оценках // Известия ТРТУ. – 2005. – № 3 (47). – С. 22-28.

Согласно просьбе автора, редакцией журнала «Известия ЮФУ. Технические науки» данная статья отозвана.

# Section IV. Report of retraction

**Glod Olga Denisovna**

**MODEL OF NETWORK PLANNING AND CONTROL IN FUZZY INTERVAL ESTIMATES**

*Glod Olga Denisovna*. Model of network planning and control in fuzzy interval estimates // Izvestiya of TRTU, 2005, No. 3 (47), pp. 22-28.

According to the author’s request, the editors of the journal "Izvestiya SFedU. Engineering Sciences" retracted the article.

## ПРАВИЛА ОФОРМЛЕНИЯ РУКОПИСЕЙ

1. Объем статьи должен быть не менее 12 и не более 18 страниц. Формат (А 4). Редактор ***Word 7 for Windows***, шрифт Times New Roman, размер 14, интервал 1,5. Авторы представляют в редакцию 1 экз. статьи и идентичный электронный вариант.

2. Названию статьи предшествует индекс УДК, соответствующий заявленной теме.

3. Текст статьи начинается с названия статьи (на русском и английском языках), фамилии, имени и отчества автора (полностью) и снабжается аннотацией на русском и английском языках объёмом ***не менее 250-300 слов***. В тексте аннотации указывается цель, задачи исследования и краткие выводы. В аннотации ***не следует*** давать ссылки на номер публикации в списке литературы к статье. После аннотаций приводятся ключевые слова (словосочетания), несущие в тексте основную смысловую нагрузку (на русском и английском языках).

4. В тексте статьи следует использовать минимальное количество таблиц и иллюстраций. Рисунок должен иметь объяснения значений всех компонентов, порядковый номер, название, расположенное под рисунком. В тексте на рисунок дается ссылка. Таблица должна иметь порядковый номер, заголовок, расположенный над ней. Данные таблиц и рисунков не должны дублировать текст. Формулы должны быть набраны ***в редакторе формул Word 7 for Windows.***

5. Цитаты тщательно сверяются с первоисточником и визируются автором на обратной стороне последней страницы: "Цитаты и фактический материал сверены". Подпись, дата.

6. Наличие пристатейного библиографического списка на русском и английском языках обязательно. ***Ссылок должно быть не менее 20-ти,*** из них на зарубежные источники – не менее 35 %. В тексте ссылки должны быть в квадратных скобках.

Примеры оформления литературы: а) для книг: фамилия, инициалы автора(ов), полное название книги, место, год издания, страницы; б) для статей: фамилия и инициалы автора(ов), полное название сборника, книги, газеты, журнала, где опубликована статья, место и год издания (сборника, книги), номер (для журнала), год и дата (для газеты), выпуск, часть (для сборника), страницы, на которых опубликована статья. Иностранная литература оформляется по тем же правилам.

Ссылки на неопубликованные работы не допускаются.

7. Рукопись должна быть тщательно вычитана. Редакционная коллегия оставляет за собой право при необходимости сокращать статьи, редактировать и отсылать авторам на доработку.

8. Статьи сопровождаются сведениями об авторе(ах) (фамилия, имя, отчество, ученое звание, должность, место работы, адрес, электронный адрес и номер телефона) на русском и английском языках.

9. Плата с аспирантов за публикацию рукописей не взымается.

**Адрес журнала в Интернете: http://izv-tn.tti.sfedu.ru/.**