

ISSN 1816-0301

# ИНФОРМАТИКА

1 (49)

ЯНВАРЬ-МАРТ  
2016

## **Редакционная коллегия:**

*Главный редактор*

**А.В. Тузиков**

*Заместитель главного редактора*

**М.Я. Ковалев**

*Члены редколлегии*

С.В. Абламейко, В.В. Анищенко, П.Н. Бибило, М.Н. Бобов,  
А.Н. Дудин, С.Я. Килин, В.В. Краснопрошин, С.П. Кундас,  
Н.А. Лиходед, П.П. Матус, С.В. Медведев, А.А. Петровский,  
Ю.Н. Сотсков, Ю.С. Харин, А.Ф. Чернявский, В.Н. Ярмолик  
Н.А. Рудая (*заведующая редакцией*)

---

*Адрес редакции:*

220012, Минск,  
ул. Сурганова, 6, к. 305  
тел. (017) 284-26-22  
e-mail: [rio@newman.bas-net.by](mailto:rio@newman.bas-net.by)  
<http://uiip.bas-net.by>

ОБЪЕДИНЕННЫЙ ИНСТИТУТ ПРОБЛЕМ ИНФОРМАТИКИ  
НАЦИОНАЛЬНОЙ АКАДЕМИИ НАУК БЕЛАРУСИ

---

---

# ИНФОРМАТИКА

---

---

ЕЖЕКВАРТАЛЬНЫЙ НАУЧНЫЙ ЖУРНАЛ

*Издается с января 2004 г.*

---

---

№ 1(49) • январь-март 2016

## СОДЕРЖАНИЕ

### МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ

- Левин Г.М., Розин Б.М., Долгий А.Б.** Оптимизация агрегирования и режимов последовательно-параллельного выполнения пересекающихся множеств операций ..... 5
- Ерофеев В.Т.** Двухсторонняя фокусировка электромагнитного поля точечного источника плоской линзой из метаматериала ..... 14
- Бондоловский А.М., Ковалев М.Я.** Оптимизация гибких цен гостиничных номеров..... 26
- Пузанов А.В.** Информационная поддержка жизненного цикла гидроприводов ..... 39

### ОБРАБОТКА СИГНАЛОВ, ИЗОБРАЖЕНИЙ И РЕЧИ

- Левчук В.А., Ковалев В.А.** Методика количественного описания биомедицинских изображений на основе словарей суперпикселей ..... 49
- Мирзаванд М.А.** Параллельная итерационная реконструкция фантома CATPHAN по экспериментальным данным ..... 58

### БИОИНФОРМАТИКА

- Новоселова Н.А., Том И.Э.** Метод построения кластеров генетических данных ..... 64
- Сергеев Р.С., Ковалев И.С., Тузиков А.В., Розенталь А., Габриэлян А.** Алгоритмы поиска мутаций лекарственной устойчивости в геномах микобактерий туберкулеза ..... 75

## ЗАЩИТА ИНФОРМАЦИИ

<b>Галибус Т.В., Краснопрошин В.В.</b> Концептуальное моделирование и организация механизмов защиты информации в распределенных системах.....	92
<b>Трубей А.И.</b> Анализ особенностей обеспечения информационной безопасности грид-систем .....	103

## КОНФЕРЕНЦИИ

<b>XV Международная конференция</b> «Развитие информатизации и государственной системы научно-технической информации» (РИНТИ–2016) .....	115
--	-----

---

---

Редактор Г.Б. Гончаренко  
Корректор А.А. Михайлова  
Компьютерная верстка О.Б. Бутевич

---

Сдано в набор 25.01.2016. Подписано в печать 01.03.2016.  
Формат 60×84 1/8. Бумага офсетная. Гарнитура Таймс. Ризография.  
Усл. печ. л. 13,4. Уч.-изд. л. 13,2. Тираж 70 экз. Заказ 1.

---

Государственное научное учреждение «Объединенный институт проблем информатики Национальной академии наук Беларуси».  
Свидетельство о государственной регистрации издателя, изготовителя, распространителя печатных изданий № 1/274 от 04.04.2014.  
ЛП № 02330/444 от 18.12.13.  
Ул. Сурганова, 6, 220012, Минск.

# INFORMATICS

---

---

PUBLISHED QUATERLY

*Issued since 2004*

---

---

№ 1(49) • January-March 2016

## CONTENTS

### MATHEMATICAL MODELING

- Levin G.M., Rozin B.M., Dolgui A.B.** Optimization of aggregation and sequential-parallel execution modes of intersecting operation sets .....5
- Erofeenko V.T.** Two-sided focusing of electromagnetic field of point source by means of planar lens from metamaterials .....14
- Bandalouski A.M., Kovalyov M.Y.** Optimizing hotel dynamic prices .....26
- Puzanov A.V.** Information support of hydrodrives lifecycle .....39

### SIGNAL, IMAGE AND SPEECH PROCESSING

- Liauchuk V.A., Kovalev V.A.** A method for quantitative description of biomedical images based on superpixel dictionaries.....49
- Mirzavand M.A.** Parallel iterative reconstruction of phantom CATPHAN on experimental data .....58

### BIOINFORMATICS

- Novoselova N.A., Tom I.E.** Method of construction of genetic data clusters .....64
- Sergeev R.S., Kavaliou I.S., Tuzikov A.V., Rosenthal A., Gabrielian A.** Algorithms for identifying drug-resistance mutations in *M. tuberculosis* genomes.....75

## INFORMATION SECURITY

<b>Galibus T., Krasnoproshin V.</b> Conceptual modeling and organization of security mechanisms in distributed systems .....	92
<b>Trubei A.I.</b> Analysis of grid-systems information security aspects.....	103

## CONFERENCES

<b>XV International conference</b> «The development of informatization and state system of scientific and technical information» (DISTI–2016) .....	115
---	-----

## МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ

УДК 658.512.2

Г.М. Левин<sup>1</sup>, Б.М. Розин<sup>1</sup>, А.Б. Долгий<sup>2</sup>**ОПТИМИЗАЦИЯ АГРЕГИРОВАНИЯ И РЕЖИМОВ  
ПОСЛЕДОВАТЕЛЬНО-ПАРАЛЛЕЛЬНОГО ВЫПОЛНЕНИЯ  
ПЕРЕСЕКАЮЩИХСЯ МНОЖЕСТВ ОПЕРАЦИЙ**

*Предложены математическая модель и метод для задачи оптимизации агрегирования и режимов последовательно-параллельного выполнения комплекса пересекающихся множеств операций. Метод основан на двухуровневой схеме декомпозиции задачи. На верхнем уровне выбирается вариант агрегирования для заданных групп операций, на нижнем уровне оптимизируются режимы выполнения операций при фиксированном варианте агрегирования.*

**Введение**

Задачам планирования процессов выполнения комплексов операций в системах различного назначения в последние десятилетия уделялось значительное внимание [1–12]. В ряде публикаций (в том числе в работах [9–12] авторов настоящей статьи) предлагаются модели и методы решения ряда задач, связанных с оптимизацией управления интенсивностью выполнения комплекса взаимосвязанных операций в предположении, что структура этого комплекса, а также структура реализующей его системы уже определены. Вместе с тем значительный научный и практический интерес представляет также разработка моделей и методов решения более сложных задач по совместной оптимизации как структуры комплекса, так и управления интенсивностями выполнения его компонентов. Под интенсивностью операции подразумевается некоторый параметр, определяющий время выполнения единицы ее объема.

В статье исследуется одна из задач совместной структурно-параметрической оптимизации при многократном последовательно-параллельном выполнении комплекса пересекающихся между собой множеств операций.

Следует отметить, что понятия «структура комплекса операций» (как и «структура реализующей его системы»), а также «управление интенсивностями (режимами) выполнения операций комплекса» могут иметь различные трактовки, причем в ряде случаев эти трактовки взаимосвязаны. В данной статье исследуется одна из ситуаций, когда структура комплекса определяется выбираемым вариантом агрегирования его операций в непересекающиеся группы (называемые в дальнейшем блоками), каждая из которых исполняется посредством своего устройства системы, причем все операции одного блока в любой момент времени выполняются с одной и той же интенсивностью. Подобная ситуация имеет место, в частности, при обработке деталей на многоинструментальных многопозиционных металлорежущих агрегатных станках, где несколько инструментов могут быть установлены в одной шпиндельной коробке (образуя единый блок технологических переходов) и иметь единую минутную подачу. Объединение операций в блоки, как правило, способствует снижению капитальных затрат на производственную систему и затрат на ее обслуживание, но одновременно приводит к возрастанию текущих материальных и временных затрат, связанных с выполнением каждой из операций комплекса в отдельности. Последнее объясняется тем, что объединение операций в блоки исключает возможность индивидуального выбора для каждой из них наилучших (с точки зрения текущих затрат) интенсивностей их выполнения.

Исследуемая задача заключается в определении варианта агрегирования операций и интенсивностей выполнения полученных блоков (включая единичные, т. е. содержащие только одну операцию), минимизирующего в совокупности стоимость выполнения комплекса работ при ограничении на общую длительность его выполнения. Рассмотренные в [9, 12] задачи являются фрагментами исследуемой, получаемыми при фиксации варианта агрегирования операций и некоторых дополнительных предположений. Наличие в исследуемой задаче комбинатор-

ной составляющей, связанной с поиском наилучшего агрегирования операций, делает эту задачу достаточно сложной, требующей специальных методов решения. Ниже предлагается один из возможных подходов к разработке таких методов.

### 1. Постановка задачи и ее математическая модель

В производственной системе планируется последовательное выполнение (однократное либо циклически повторяющееся) всех работ комплекса, образованного множеством  $I = \{1, \dots, i, \dots, n\}$  различных работ, причем этот комплекс включает  $n_i$  идентичных работ  $i \in I$ . Выполнение очередной работы  $i \in I$  комплекса заключается в параллельном выполнении всех операций соответствующего подмножества  $J_i$  исходного множества  $J = \{1, \dots, m\}$  операций. Подмножества семейства  $\mathbf{J} = \{J_1, \dots, J_i, \dots, J_n\}$ , образующие работы комплекса, могут пересекаться и  $\bigcup_{i=1}^n J_i = J$ .

Задано некоторое семейство  $\mathbf{W}$  непересекающихся неединичных подмножеств множества  $J$ , каждое из которых является потенциальным блоком операций, причем любое подмножество  $w \in \mathbf{W}$  может содержать операции из разных подмножеств семейства  $\mathbf{J}$  (рис. 1). Подмножества  $w \in \mathbf{W}$  выделены на рис. 1 штриховыми линиями.

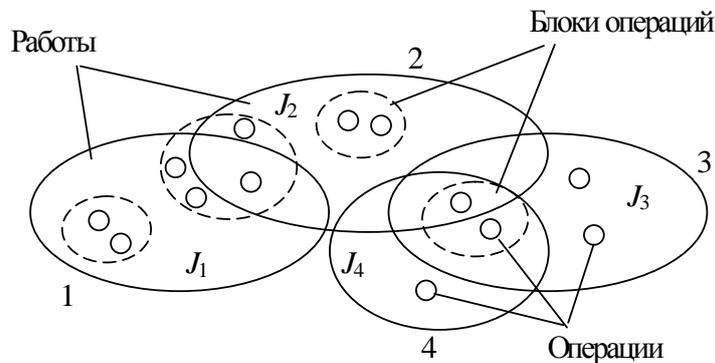


Рис. 1. Пример комплекса работ

Исследуется один из частных случаев возможного агрегирования операций, когда каждое из множеств операций  $w \in \mathbf{W}$  может выполняться либо только полностью агрегированным (т. е. как один блок), либо полностью разагрегированным (т. е. каждая операция автономна). В этом случае множество возможных вариантов агрегирования представимо множеством  $\mathcal{Q}$  всевозможных значений двоичного вектора  $q = (q_w | w \in \mathbf{W})$ , где  $q_w = 0$  при агрегированном выполнении операций множества  $w \in \mathbf{W}$  и  $q_w = 1$  в противном случае.

Если операции из  $w \in \mathbf{W}$  выполняются единым блоком (т. е. агрегированно), то их интенсивности одинаковы. При автономном выполнении этих операций каждая из них может иметь свою отличную от других операций из  $w$  интенсивность. В дальнейшем каждая автономно выполняемая операция также рассматривается как блок (единичный блок). Здесь и далее под интенсивностью  $s_j$  операции  $j \in J$  подразумевается время, затрачиваемое на выполнение единицы ее объема. Для каждой операции  $j \in J$  известны диапазон  $S_j = [s_{1j}, s_{2j}]$  возможных значений ее интенсивности  $s_j$  и объем  $V_{ij}$  в содержащей ее работе  $i \in I$ . Естественно предположить, что для любого блока  $w \in \mathbf{W}$  диапазон возможных значений интенсивностей выполнения его операций  $S_w = [S_{1w} = \max\{s_{1j} | j \in w\}, S_{2w} = \min\{s_{2j} | j \in w\}] \neq \emptyset$ , поскольку в противном случае все операции из соответствующего  $w$  не могут быть агрегированы в одном блоке и, следовательно, это множество должно быть исключено из  $\mathbf{W}$ .

Ограничимся случаем, когда интенсивность выполнения каждой операции (а значит, и блока операций) выбирается однократно, не зависит от работы, в составе которой эта операция выполняется, а также не изменяется во времени. Таким образом, если  $s = (s_j | j \in J)$  – вектор принимаемых значений интенсивностей операций  $j \in J$  и  $j(w)$  – некоторая фиксированная операция из  $w \in \mathbf{W}$ , то дли-

тельность  $t_{ij}(s_j)$  операции  $j$  в составе работы  $i$  равна  $V_{ij}s_j$ , длительность  $t_{iw}(s_{j(w)})$  агрегированного выполнения блока  $w$  в составе работы  $i$  –  $\max\{V_{ij}|j \in J_i \cap w\}s_{j(w)}$ , а длительность  $t_i(s)$  работы  $i$  –  $\max\{t_{ij}(s_j) | j \in J_i\}$ .

Анализ реальных ситуаций показывает, что для многих из них с достаточной для практики точностью можно считать, что общие материальные и временные затраты на выполнение каждой из работ комплекса складываются из двух основных частей. Первой является сумма затрат, связанных с восстановлением ресурсов, расходуемых при выполнении каждой из операций работы. Количество соответствующего ресурса, расходуемого на операцию  $j \in J$ , обычно зависит от работы  $i$ , в составе которой она выполняется, объема  $V_{ij}$  операции в этой работе и принятой интенсивности  $s_j$  операции. Максимальное количество каждого ресурса, которое может быть в системе, ограничено. Полное восстановление любого из ресурсов осуществляется после его расходования до допустимого уровня по завершении работы, в которой это произошло. Выполнение очередной работы может начаться лишь после завершения процесса восстановления соответствующего ресурса. Материальные и временные затраты на восстановление каждого из ресурсов, как правило, известны. В этом случае можно считать заданными определенные на  $S_j$  невозрастающие положительные функции  $f_{1ij}(s_j)$  и  $f_{2ij}(s_j)$ , представляющие зависимости удельных (отнесенных к единице объема) соответственно материальных и временных затрат на выполнение операции  $j$  в составе работы  $i$  от принимаемой интенсивности  $s_j$  этой операции.

Вторую часть представляют материальные и временные затраты, связанные с амортизацией оборудования и его обслуживанием. Эти затраты можно считать пропорциональными длительности  $t_i(s)$  работы  $i$  при принятых интенсивностях составляющих эту работу операций. Удельные (отнесенные к единице длительности работы) затраты этой категории зачастую различны для разных работ  $i \in I$  и зависят от принимаемого варианта  $q \in Q$  агрегирования операций, причем во многих реальных ситуациях достаточно предположить, что для всех работ  $i \in I$  эти зависимости  $C_{pi}(q)$  имеют следующую структуру:

$$C_{pi}(q) = E_{pi} + \sum_{w \in \mathbf{W}} e_{piw} q_w,$$

где  $p = 1$  или  $p = 2$  при оценке материальных или временных затрат соответственно;  $E_{pi} > 0$  – величина удельных затрат на работу  $i \in I$  в случае агрегированного выполнения операций для всех множеств  $w \in \mathbf{W}$ ;  $e_{piw} > 0$  – дополнительные удельные затраты на работу  $i \in I$  при автономном выполнении операций множества  $w \in \mathbf{W}$ . В дальнейшем параметры  $E_{pi}$  и  $e_{piw}$  предполагаются известными для всех  $i \in I$ ,  $w \in \mathbf{W}$  и  $p = 1, 2$ .

Пусть  $S = \prod_{j \in J} S_j$ ;  $q \in Q$  и  $S(q)$  – множество таких векторов  $s \in S$ , что для любых  $w \in \mathbf{W}$  и  $j \in w$

значение  $s_j = s_{j(w)}$  при  $q_w = 0$ . Заметим, что в силу оговоренного ранее  $S(q) \neq \emptyset$  для любых  $q \in Q$ .

Согласно принятым предположениям о специфике рассматриваемого комплекса операций и его свойствах при фиксированных значениях векторов  $q \in Q$  и  $s \in S(q)$ , представляющих вариант агрегирования операций и их интенсивности, общие средние материальные ( $p = 1$ ) и временные ( $p = 2$ ) затраты на однократное выполнение всего комплекса работ имеют следующий вид:

$$F_p(q, s) = \sum_{i \in I} n_i [(E_{pi} + \sum_{w \in \mathbf{W}} e_{piw} q_w) \max\{V_{ij}s_j | j \in J_i\} + \sum_{j \in J_i} V_{ij} f_{pij}(s_j)].$$

Пусть  $T^0$  – максимально допустимое значение общих затрат времени на однократное выполнение всех работ комплекса, определяемое исходя из заданной производительности. Тогда исследуемая задача определения варианта агрегирования операций и их интенсивностей, минимизирующая общие материальные затраты на однократное выполнение всего комплекса операций при обеспечении требуемой производительности, сводится к следующей задаче смешанного нелинейного программирования:

$$F_1(q, s) \rightarrow \min; \quad (1)$$

$$F_2(q, s) \leq T^0; \quad (2)$$

$$q \in Q; \tag{3}$$

$$s \in S(q). \tag{4}$$

В ряде публикаций (например, [2, 3, 6]) описываются производственные системы, планирование функционирования которых включает решение подобных задач.

## 2. Метод решения

Отметим следующие особенности задачи (1)–(4) (называемой далее задачей **A**), которые могут быть использованы при выборе подходов к ее решению:

- наличие двух групп разнородных переменных ( $|\mathbf{W}|$ -мерного двоичного вектора  $q \in Q$  и  $m$ -мерного вектора  $s$  с компонентами, принимающими значения из заданных отрезков);
- одинаковая структура функций  $F_1(q, s)$  и  $F_2(q, s)$ ;
- если векторы  $q', q'' \in Q$  таковы, что  $q'_w \leq q''_w$  для всех  $w \in \mathbf{W}$ , то  $S(q') \subseteq S(q'')$  и  $F_p(q', s) \leq F_p(q'', s)$ ; если при этом  $q' \neq q''$ , то  $F_p(q', s) < F_p(q'', s)$  для любых  $s \in S(q')$  и  $p = 1, 2$ .

Для решения задачи **A** можно воспользоваться следующей двухуровневой схемой (рис. 2). На нижнем уровне решается подзадача **B**( $q$ ) нелинейного программирования по определению для фиксированного значения  $q \in Q$  такого значения  $s^*(q)$  вектора  $s \in S(q)$ , которое минимизирует функцию  $F_1(q, s)$  при условии, что  $F_2(q, s) \leq T^0$ . На верхнем уровне решается подзадача **C** дискретного программирования по поиску такого значения  $q^*$  вектора  $q \in Q$ , которое минимизирует функцию  $\Phi(q) = F_1(q, s^*(q))$ , при этом предполагается, что  $\Phi(q) = \infty$ , если подзадача **B**( $q$ ) не имеет решения.

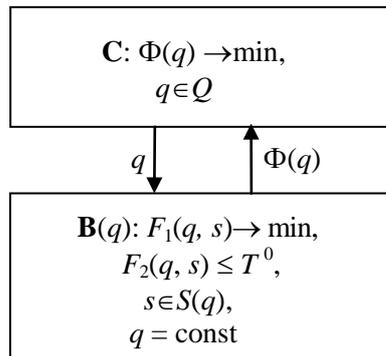


Рис. 2. Схема декомпозиции задачи **A**

Поскольку  $q^*$  и  $s^*(q)$  являются решениями подзадач **C** и **B**( $q$ ) соответственно, то  $(q^*, s^*) = (q^*, s^*(q^*))$  – решение исходной задачи **A**. Если же  $q^*$  и  $s^*(q)$  являются лишь приближенными решениями соответствующих подзадач, то вектор  $(q^*, s^*)$  может быть принят в качестве приближенного решения задачи **A**.

Отметим некоторые из возможных подходов к решению выделенных подзадач. Для более детального рассмотрения подзадачи **B**( $q$ ) нижнего уровня введем некоторые дополнительные обозначения, опуская при этом фиксируемое в подзадаче значение вектора  $q$ .

Пусть  $I_j = \{i \in I | j \in J_i\}$  – множество работ, содержащих операцию  $j \in J$ . Для данного фиксированного варианта агрегирования операций положим:  $\mathbf{W}^0 = \{w \in \mathbf{W} | q_w = 0\}$  – множество реализуемых в этом варианте блоков;  $\mathbf{W}_i^0 = \{w \in \mathbf{W}^0 | w \cap J_i \neq \emptyset\}$  – множество блоков, реализуемых в работе  $i \in I$ ;  $J^1 = \{j \in J | j \notin w \in \mathbf{W}^0\}$  – множество операций, реализуемых в этом варианте автономно;  $J_i^1 = J_i \cap J^1$  – множество операций, реализуемых автономно в работе  $i \in I$ ;  $\tilde{E}_{pi} = n_i (E_{pi} + \sum_{w \in \mathbf{W} \setminus \mathbf{W}^0} e_{piw})$ ;

$$\bar{V}_{iw} = \max\{V_{ij} | j \in J_i \cap w\}; \quad \tilde{f}_{pj}(s_j) = \sum_{i \in I_j} n_i V_{ij} f_{pij}(s_j); \quad \tilde{f}_{pw}(s_w) = \sum_{j \in w} \tilde{f}_{pj}(s_w), \quad w \in \mathbf{W}^0; \quad \bar{s} = ((s_j | j \in J^1), (s_w | w \in \mathbf{W}^0))$$

и  $\bar{S} = \prod_{j \in J^1} S_j \times \prod_{w \in \mathbf{W}^0} S_w$ . В этих обозначениях функции  $F_p(q, s)$ ,  $p = 1, 2$ , представимы функциями

$$\tilde{F}_p(\bar{s}) = \sum_{i \in I} \tilde{E}_{pi} \max\{\max\{V_{ij}s_j \mid j \in J_i^1\}, \max\{\bar{V}_{iw}s_w \mid w \in \mathbf{W}_i^0\}\} + \sum_{j \in J^1} \tilde{f}_{pj}(s_j) + \sum_{w \in \mathbf{W}^0} \tilde{f}_{pw}(s_w), p = 1, 2,$$

и задача  $\mathbf{B}(q)$  при фиксированном значении  $q \in Q$  заключается в определении такого значения вектора  $\bar{s} \in \bar{S}$ , которое минимизирует функцию  $\tilde{F}_1(\bar{s})$  при условии, что  $\tilde{F}_2(\bar{s}) \leq T^0$ .

Возможные подходы к решению задачи  $\mathbf{B}(q)$  во многом зависят от свойств функций  $f_{pij}(\bullet)$  и соотношений объемов  $V_{ij}$  операций  $j$  в различных содержащих их работах  $i \in I$  и потенциальных блоках операций  $w \in \mathbf{W}$ . Выделим следующих два распространенных на практике частных случая: а) когда функции  $f_{pij}(s_j)$  выпуклы для всех  $p = 1, 2, i \in I$  и  $j \in J_i$  (задача  $\mathbf{B}_1(q)$ ); б) когда для каждой операции  $j \in J^1$  ее объемы  $V_{ij}$  одинаковы и равны  $V_j$  для всех работ  $i \in I_j$  и когда для каждого блока  $w \in \mathbf{W}^0$  значения  $\bar{V}_{iw}$  одинаковы и равны  $\bar{V}_w$  для всех таких работ  $i \in I$ , что  $w \cap J_i \neq \emptyset$  (задача  $\mathbf{B}_2(q)$ ). В остальных случаях разработка эффективных методов решения подзадачи  $\mathbf{B}(q)$  требует дополнительных исследований.

В задаче  $\mathbf{B}_1(q)$  при принятых предположениях функции  $\tilde{f}_{pj}(s_j)$  и  $\tilde{f}_{pw}(s_w)$  являются выпуклыми на отрезках  $S_j$  и  $S_w$  соответственно для всех  $j \in J^1$  и  $w \in \mathbf{W}^0$ , и для ее решения применимы известные методы выпуклого программирования. В силу сепарабельности последних двух слагаемых в функциях  $\tilde{F}_1(\bar{s})$  и  $\tilde{F}_2(\bar{s})$  для решения задачи  $\mathbf{B}_1(q)$  может быть адаптирован, в частности, подход, предложенный в [12] для решения аналогичной задачи и основанный на ее аппроксимации задачей линейного программирования.

Построим кусочно-линейные аппроксимации функций  $\tilde{f}_{pj}(s_j)$  и  $\tilde{f}_{pw}(s_w)$  на отрезках  $S_j$  и  $S_w$  соответственно, полагая

$$\tilde{f}_{pj}(s_j) \approx \max\{a_{pjk}s_j + b_{pjk} \mid k = 1, \dots, r_{pj}\}, p = 1, 2, j \in J; \quad (5)$$

$$\tilde{f}_{pw}(s_w) \approx \max\{c_{pwk}s_w + d_{pwk} \mid k = 1, \dots, u_{pw}\}, p = 1, 2, w \in \mathbf{W}, \quad (6)$$

где  $a_{pjk}, b_{pjk}, c_{pwk}, d_{pwk}, r_{pj}$  и  $u_{pw}$  являются параметрами этой аппроксимации, причем параметры  $r_{pj}$  и  $u_{pw}$  во многом определяют ее точность. Заметим, что параметры аппроксимации не зависят от значения вектора  $q$  и, следовательно, эти аппроксимации могут быть построены заранее до непосредственного решения исходной задачи  $\mathbf{A}$  и использоваться в задачах  $\mathbf{B}_1(q)$  при различных значениях вектора  $q$ .

Тогда приближенное решение задачи  $\mathbf{B}_1(q)$  при фиксированном  $q$  может быть получено в результате решения следующей задачи линейного программирования:

$$\sum_{i \in I} \tilde{E}_{1i} t_i + \sum_{j \in J^1} y_{1j} + \sum_{w \in \mathbf{W}^0} z_{1w} \rightarrow \min; \quad (7)$$

$$\sum_{i \in I} \tilde{E}_{2i} t_i + \sum_{j \in J^1} y_{2j} + \sum_{w \in \mathbf{W}^0} z_{2w} \leq T^0; \quad (8)$$

$$t_i - V_{ij}s_j \geq 0, \quad i \in I, \quad j \in J_i^1; \quad (9)$$

$$t_i - \bar{V}_{iw}s_w \geq 0, \quad i \in I, \quad w \in \mathbf{W}_i^0; \quad (10)$$

$$y_{pj} - a_{pjk}s_j \geq b_{pjk}, \quad p = 1, 2, \quad j \in J^1, \quad k = 1, \dots, r_{pj}; \quad (11)$$

$$z_{pw} - c_{pwk}s_w \geq d_{pwk}, \quad p = 1, 2, \quad w \in \mathbf{W}^0, \quad k = 1, \dots, u_{pw}; \quad (12)$$

$$s_j \in S_j, \quad j \in J^1; \quad (13)$$

$$s_w \in S_w, \quad w \in \mathbf{W}^0. \quad (14)$$

Искомыми в этой задаче являются векторы  $t = (t_i \mid i \in I)$ ,  $y = (y_{pj} \mid p = 1, 2, j \in J^1)$ ,  $z = (z_{pw} \mid p = 1, 2, w \in \mathbf{W}^0)$  и  $\bar{s} \in \bar{S}$ . Если  $(t^*, y^*, z^*, \bar{s}^*)$  – ее решение, то вектор  $s$  с компонентами  $s_j = s_j^*$  при  $j \in w \in \mathbf{W}^0$  и  $s_j = s_j^*$

при  $j \in J^1$  может быть принят в качестве приближенного решения  $s^*(q)$  задачи  $\mathbf{B}_1(q)$ . Несовпадение минимальных значений целевых функций этих задач определяется точностью аппроксимаций (5) – (6) функций  $\tilde{f}_{pj}(s_j)$  и  $\tilde{f}_{pw}(s_w)$  в окрестности решения исходной задачи  $\mathbf{B}_1(q)$ .

Для решения аппроксимирующих задач (7) – (14) при фиксированных значениях  $q \in Q$  могут быть использованы существующие программные средства типа CPLEX.

В задаче  $\mathbf{B}_2(q)$  в соответствии со сделанными предположениями функции  $\tilde{F}_1(\bar{s})$ ,  $\tilde{F}_2(\bar{s})$  имеют вид

$$\tilde{F}_p(\bar{s}) = \sum_{i \in I} \tilde{E}_{pi} \max \{ \max \{ V_j s_j \mid j \in J_i^1 \}, \max \{ \bar{V}_w s_w \mid w \in W_i^0 \} \} + \sum_{j \in J^1} \tilde{f}_{pj}(s_j) + \sum_{w \in W^0} \tilde{f}_{pw}(s_w), p = 1, 2.$$

Для ее решения может быть применен подход, аналогичный использованному в [9] для решения подобной задачи. Этот подход основан на комбинации методов множителей Лагранжа и динамического программирования.

Введем в рассмотрение функцию Лагранжа

$$L(\lambda, \bar{s}) = \lambda \tilde{F}_1(\bar{s}) + (1 - \lambda) \tilde{F}_2(\bar{s}) = \sum_{i \in I} \tilde{E}_i(\lambda) \max \{ \max \{ V_j s_j \mid j \in J_i^1 \}, \max \{ \bar{V}_w s_w \mid w \in W_i^0 \} \} + \\ + \sum_{j \in J^1} \varphi_{1j}(\lambda, s_j) + \sum_{w \in W^0} \varphi_{2w}(\lambda, s_w)$$

с множителем  $\lambda \in [0, 1]$  и подзадачу  $\mathbf{D}(\lambda)$  минимизации этой функции по  $\bar{s} \in \bar{S}$  при фиксированном значении  $\lambda \in [0, 1]$ . Здесь и далее параметр  $q$  по-прежнему опущен,  $\tilde{E}_i(\lambda) = \lambda \tilde{E}_{1i} + (1 - \lambda) \tilde{E}_{2i}$ ,  $\varphi_{1j}(\lambda, s_j) = \lambda \tilde{f}_{1j}(s_j) + (1 - \lambda) \tilde{f}_{2j}(s_j)$  и  $\varphi_{2w}(\lambda, s_w) = \lambda \tilde{f}_{1w}(s_w) + (1 - \lambda) \tilde{f}_{2w}(s_w)$ .

Пусть  $\bar{s}^*(\lambda)$  – решение этой задачи при некотором значении  $\lambda \in [0, 1]$  и  $H(\lambda) = \tilde{F}_2(\bar{s}^*(\lambda))$ . Очевидно, что если  $H(0) > T^0$ , то задача  $\mathbf{B}_2(q)$  неразрешима (и, следовательно, соответствующее  $\Phi(q) = \infty$ ), а если  $H(1) \leq T^0$ , то  $\bar{s}^*(1)$  – решение этой задачи. Поскольку функция  $H(\lambda)$  не убывает на отрезке  $[0, 1]$ , то иначе решение задачи  $\mathbf{B}_2(q)$  может быть получено по двухуровневой схеме. На нижнем уровне для фиксированного значения  $\lambda \in (0, 1)$  решается подзадача  $\mathbf{D}(\lambda)$  по определению  $\bar{s}^*(\lambda)$ , на верхнем отыскивается наибольший корень  $\lambda^* \in (0, 1)$  уравнения  $H(\lambda) = T^0$  с монотонной левой частью (подзадача  $\mathbf{L}$ ).

Если множество  $\{(\tilde{F}_1(\bar{s}), \tilde{F}_2(\bar{s})) \mid \bar{s} \in \bar{S}\}$  эффеktivно выпукло (см., в частности, [13]), то  $\bar{s}^*(\lambda^*)$  – точное решение задачи  $\mathbf{B}_2(q)$ , в противном случае  $\bar{s}^*(\lambda^*)$  может быть принято в качестве ее приближенного решения.

Решение подзадачи  $\mathbf{L}$  может быть получено известными методами, обычно используемыми для нахождения корней уравнений с монотонной левой частью, а решение задачи  $\mathbf{D}(\lambda)$  при фиксированном  $\lambda$  может быть сведено к отысканию кратчайшего пути в некотором бесконтурном орграфе  $\mathbf{G}$ . Для более детального представления идеи этого сведения введем следующие определения и обозначения:

$$M_1(I') = \bigcup_{i \in I'} J_i^1, \quad M_2(I') = \bigcup_{i \in I'} W_i^0 \quad \text{и} \quad M(I') = (M_1(I'), M_2(I')) \quad \text{для} \quad I' \subseteq I;$$

$\Omega$  – семейство подмножеств  $I' \subseteq I$ , для которых не существует другого подмножества  $I'' \subseteq I$ , что  $I' \subset I''$  и  $M(I') = M(I'')$ . Предполагается, что  $\Omega$  включает также и пустое подмножество;

$\tilde{\Omega}$  – множество таких пар  $(I^1, I^2) \in \Omega \times \Omega$ , что  $I^1 \subset I^2$  и не существует такого  $I^3 \in \Omega$ , что  $I^1 \subset I^3 \subset I^2$ ;

$$N_k(I', I'') = M_k(I'') \setminus M_k(I'), \quad k = 1, 2, \quad \text{для} \quad (I', I'') \in \tilde{\Omega};$$

$$\tau_{1k}(J) = \max \{ V_j s_{kj} \mid j \in J \} \quad \text{и} \quad \tau_{2k}(W) = \max \{ \bar{V}_w s_{kw} \mid w \in W \}, \quad J' \subseteq J, \quad W' \subseteq W \quad \text{и} \quad k = 1, 2;$$

$$s_j(\tau) = \min[s_{2j}, \tau/V_j], j \in \mathbf{J}, \text{ и } \underline{s}_w(\tau) = \min[S_{2w}, \tau/\bar{V}_w], w \in \mathbf{W};$$

$$\psi(\lambda, I', J', W', \tau) = \sum_{i \in I'} \tilde{E}_i(\lambda)\tau + \sum_{j \in J'} \phi_{1j}(\lambda, s_j(\tau)) + \sum_{w \in W'} \phi_{2w}(\lambda, \underline{s}_w(\tau)), I' \subseteq I, J' \subseteq J, W' \subseteq \mathbf{W}$$

и  $\tau \geq \max[\tau_{11}(J), \tau_{21}(W)];$

$\bar{\tau}(\lambda, I', J', W', \tau)$  – значение  $\tau \in [\max\{\tau', \tau_{11}(J), \tau_{21}(W)\}, \max\{\tau', \tau_{12}(J), \tau_{22}(W)\}]$ , которому соответствует наименьшее значение функции  $\psi(\lambda, I', J', W', \tau)$ , где  $\tau' > 0$ .

Очевидно, что  $M_k(I) \subset M_k(I'')$  по крайней мере для одного из  $k = 1, 2$  для любой пары  $(I', I'') \in \tilde{\Omega}$ .

Построение орграфа  $\mathbf{G}$  возможно по следующей рекуррентной схеме:

1. Пара  $(\emptyset, 0)$  представляет начальную вершину этого орграфа.

2. Если вершина  $(I', \tau')$  принадлежит  $\mathbf{G}$ , то дуга  $((I', \tau'), (I'', \tau''))$  также принадлежит  $\mathbf{G}$ , если  $(I', I'') \in \tilde{\Omega}$  и  $\tau'' \in [\max\{\tau', \tau_{11}(N_1(I', I'')), \tau_{21}(N_2(I', I''))\}, \bar{\tau}(\lambda, I'' \setminus I', N_1(I', I''), N_2(I', I''), \tau')]$ . Длина этой дуги принимается равной  $\psi(\lambda, I'' \setminus I', N_1(I', I''), N_2(I', I''), \tau')$ .

Следует заметить, что параметр  $\tau'$  в парах  $(I', \tau')$  представляет максимальное из предполагаемых времен выполнения работ из множества  $I'$ . Размерность орграфа  $\mathbf{G}$  определяется как мощностью множества  $\tilde{\Omega}$ , так и длинами отрезков возможных значений параметра  $\tau'$  в парах  $(I', \tau')$  и принимаемой при построении орграфа дискретизацией этих отрезков.

Следуя [9], можно показать, что задача  $\mathbf{D}(\lambda)$  сводится к нахождению в орграфе  $\mathbf{G}$  кратчайшего пути из начальной вершины  $(\emptyset, 0)$  в множество конечных вершин вида  $(I=I, \tau)$ . Если дуга  $((I', \tau'), (I'', \tau''))$  принадлежит такому пути, то в решении  $\bar{s}^*(\lambda)$  задачи  $\mathbf{D}(\lambda)$  принимается значение  $s_j^*(\lambda) = s_j(\tau')$  для всех  $j \in N_1(I', I'')$  и  $s_w^*(\lambda) = s_w(\tau'')$  для всех  $w \in N_2(I', I'')$ .

Пусть  $H(I', \tau)$  – длина кратчайшего пути в орграфе  $\mathbf{G}$  из его начальной вершины  $(\emptyset, 0)$  в вершину  $(I', \tau)$ . При построении орграфа  $\mathbf{G}$  и поиске в нем кратчайшего пути в множество конечных вершин промежуточная вершина  $(I', \tau')$  может быть исключена из дальнейшего рассмотрения, если определится другая такая вершина  $(I'', \tau'')$ , что  $I' \subseteq I'', \tau' \geq \tau''$  и  $H(I', \tau) \geq H(I'', \tau'')$ .

В задаче  $\mathbf{C}$  верхнего уровня число возможных значений вектора  $q$  равно  $2^{|\mathbf{W}|}$ , поэтому ее решение полным перебором всего множества  $\mathcal{Q}$  требует значительных затрат времени даже при сравнительно небольших значениях  $|\mathbf{W}|$  и практически нереализуемо при больших  $|\mathbf{W}|$ . Для сокращения перебора могут быть использованы известные методы, основанные на идеях случайного поиска, эвристиках и метаэвристиках в сочетании с предлагаемым ниже вариантом метода последовательной фиксации переменных (МПФП).

Пусть имеется некоторое значение  $q^0$  вектора  $q \in \mathcal{Q}$ , полученное с использованием перечисленных выше эвристических подходов. Для последующей оптимизации  $q^0$  предлагается следующий алгоритм МПФП. Алгоритм сводится к пошаговой фиксации значений (0 либо 1) одной из незафиксированных на предыдущих шагах компонент вектора  $q^0$ . Число оцениваемых векторов  $q \in \mathcal{Q}$  на очередном шаге алгоритма последовательно сокращается от  $|\mathbf{W}|$  до 1, при этом среди возможных вариантов выбирается единственный с наименьшим значением целевой функции.

Ниже для описания итерации алгоритма используются следующие обозначения:

$\Psi$  – список блоков  $w \in \mathbf{W}$  с фиксированными в дальнейшем значениями  $q_w$  для текущего вектора  $q^{mek}$ ;

$\underline{q}(q, w)$  – вектор из множества  $\mathcal{Q}$ , который отличается от вектора  $q \in \mathcal{Q}$  только компонентой  $q_w$ .

*Итерация алгоритма МПФП:*

1. Положим  $q^{mek} = q^0$ ;  $L_{mek} = \Phi(q^0)$ ,  $\Psi = \emptyset$ .

2. Если  $\Psi = \mathbf{W}$ , то переход к п. 4. Иначе переход к п. 3.

3. Посредством решения серии задач  $\mathbf{B}(q)$  находится  $w^* = \operatorname{argmin}\{\Phi(\underline{q}(q, w)) | w \in \mathbf{W} \setminus \Psi\}$ . Если  $L_{mek} \leq L^* = \Phi(\underline{q}(q, w^*))$ , то переход к п. 4. Иначе полагается  $L_{mek} = L^*$ ,  $q^{mek} = \underline{q}(q, w^*)$ ,  $\Psi = \Psi \cup \{w^*\}$  и переход к п. 2.

4. Вектор  $q^{mek}$  принимается в качестве текущего решения  $q^0$  задачи  $\mathbf{C}$ . Конец итерации.

Итерация алгоритма МПФП может повторяться для текущего решения  $q^0$ , пока функция  $\Phi(q^0)$  убывает. Число задач  $\mathbf{B}(q)$ , решаемых на каждой итерации алгоритма, не превышает  $O(|\mathbf{W}|^2)$ .

Одно из возможных направлений развития алгоритма МПФП связано с проверкой целесообразности изменения одновременно нескольких компонент вектора  $q^{mek}$ , отбираемых по результатам решения задач  $\mathbf{B}(q)$  в п. 3.

*Замечание.* В некоторых случаях заранее известно, что оптимальные интенсивности  $s_j^*(q_w)$  всех операций  $j \in w$  некоторого блока  $w \in \mathbf{W}$  одинаковы (например, когда операции  $j \in w$  идентичны). Тогда для такого блока можно заранее положить  $q_w^* = 0$  в решении  $q^*$  задачи С. Это свойство позволяет сократить размерность данной задачи.

### Заключение

Предложены математическая модель и метод решения задачи комплексной оптимизации агрегирования в блоки и интенсивностей последовательно-параллельного выполнения комплекса пересекающихся множеств операций. Рассмотрены частные случаи задачи, когда интенсивность любой операции остается неизменной в составе любого множества, при этом зависимости материальных и временных затрат на выполнение любой операции в составе конкретного множества от ее интенсивности представимы выпуклыми функциями, либо объемы любой операции (блока операций) в составе различных включающих ее множеств одинаковы. Выбор степени агрегирования заранее определенных подмножеств операций осуществляется из двух крайних вариантов: либо все операции, входящие в состав подмножества, выполняются одним общим устройством и имеют общую интенсивность, либо интенсивности всех операций такого подмножества выбираются индивидуально.

Описана двухуровневая декомпозиционная схема решения рассматриваемой задачи. Метод приближенного решения задачи верхнего уровня основан на комбинации эвристических методов и методе последовательной фиксации переменных. Для подзадачи нижнего уровня поиска оптимальных интенсивностей операций при фиксированном варианте агрегирования предложены методы решения для двух частных случаев исходной задачи: комбинация методов Лагранжевой релаксации и динамического программирования для случая равных объемов любой операции (блока операций) в разных работах и метод аппроксимации задачей линейного программирования для случая выпуклых функций затрат на операции.

В качестве направления дальнейших исследований предполагается рассмотреть новые постановки задач планирования структуры управления и интенсивности выполнения комплексов операций, в которых:

- варианты агрегирования операций в составе заданных подмножеств операций определяются различными комбинациями подмножеств этих операций;
- интенсивности выполнения операций могут зависеть от работы, в составе которой эти операции выполняются;
- функции затрат на выполнение операций не являются выпуклыми и объемы операций (блоков операций) различны для разных работ.

### Список литературы

1. Alting, L. Computer Aided Process Planning: the state-of-the-art survey / L. Alting, H. Zhang // Int. J. Prod. Res. – 1989. – Vol. 27, № 4. – P. 553–585.
2. Halevi, G. Process and Operation Panning / G. Halevi. – Springer, 2003. – 335 p.
3. Bukchin, J. Design of flexible assembly line to minimize equipment cost / J. Bukchin, M. Tzur // IE Transactions. – 2000. – Vol. 32. – P. 585–598.
4. Gupta, A.K. Optimization of due-date objectives in scheduling semiconductor batch manufacturing / A.K. Gupta, A.I. Sivakumar // Intern. J. of Machine Tools and Manufacture. – 2006. – Vol. 46. – P. 1671–1679.
5. Burkov, V.N. Models and methods of multiprojects' management / V.N. Burkov, D.A. Novikov // Systems Science. – 1999. – Vol. 256, № 2. – P. 5–14.

6. Dolgui, A. Graph approach for optimal design of transfer machine with rotary table / A. Dolgui, N. Guschinsky, G. Levin // Intern. J. of Production Research. – 2009. – Vol. 47, № 2. – P. 321–341.
7. Dolgui, A. Enhanced mixed integer programming model for a transfer line design problem / A. Dolgui, N. Guschinsky, G. Levin // Computers and Industrial Engineering. – 2012. – Vol. 62, № 2. – P. 570–578.
8. Левин, Г.М. Оптимизация режимов параллельной многоинструментальной обработки деталей на агрегатном оборудовании с учетом групповой смены инструментов / Г.М. Левин, Б.М. Розин // Информатика. – 2011. – № 3. – С. 33–47.
9. Левин, Г.М. Оптимизация последовательно-параллельного выполнения комплекса взаимосвязанных операций / Г.М. Левин, Б.М. Розин // Весці НАН Беларусі. Сер. фіз.-мат. навук. – 2013. – № 1. – С. 111–116.
10. Levin, G. Optimization of Multi-tool Cutting Modes for Batch Manufacturing in Large Series Machining Environment / G. Levin, B. Rozin, A. Dolgui // Proc. of the 14th IFAC Symp. on Information Control Problems in Manufacturing (INCOM'12), Bucharest, Romania, May 23–25, 2012. – Elsevier Science, 2012. – P. 444–448.
11. Rozin, B. Optimization of Multi-tool Cutting Modes in Multi-item Batch Manufacturing System / B. Rozin, G. Levin, A. Dolgui // Proc. of the IFAC Conf. on Manufacturing Modelling, Management and Control (MIM'2013), Russia, June 19–21, 2013. – Elsevier Science, 2013. – P. 766–771.
12. Левин, Г.М. Линейная аппроксимация задачи оптимизации интенсивностей последовательно-параллельного выполнения пересекающихся множеств операций / Г.М. Левин, Б.М. Розин, А.Б. Долгий // Информатика. – 2014. – № 3. – С. 44–51.
13. Подиновский, В.В. Парето-оптимальные решения многокритериальных задач / В.В. Подиновский, В.Д. Ногин. – М. : Наука. Глав. ред. физ.-мат. лит-ры, 1982. – 256 с.

Поступила 24.11.2015

<sup>1</sup>Объединенный институт проблем информатики НАН Беларуси,  
Минск, Сурганова, 6  
e-mail: {levin; rozin}@newman.bas-net.by

<sup>2</sup>Ecole Nationale Supérieure des Mines, CNRS,  
UMR6597 IRCCYN, F-44307 Nantes  
Cedex 3, France  
e-mail : alexandre.dolgui@mines-nantes.fr

**G.M. Levin, B.M. Rozin, A.B. Dolgui**

## **OPTIMIZATION OF AGGREGATION AND SEQUENTIAL-PARALLEL EXECUTION MODES OF INTERSECTING OPERATION SETS**

A mathematical model and a method for the problem of optimization of aggregation and of sequential-parallel execution modes of intersecting operation sets are proposed. The proposed method is based on the two-level decomposition scheme. At the top level the variant of aggregation for groups of operations is selected, and at the lower level the execution modes of operations are optimized for a fixed version of aggregation.

УДК 537.874:621.372.8

В.Т. Ерофеевко

## ДВУХСТОРОННЯЯ ФОКУСИРОВКА ЭЛЕКТРОМАГНИТНОГО ПОЛЯ ТОЧЕЧНОГО ИСТОЧНИКА ПЛОСКОЙ ЛИНЗОЙ ИЗ МЕТАМАТЕРИАЛА

*Предлагается аналитическое решение краевой задачи проникновения электромагнитного поля диполя через плоскую линзу из фокусирующих материалов. Показывается, что для линз из фокусирующих материалов образуются два фокуса: один за экраном и второй перед экраном в области источника. Поле фокуса перед экраном равняется сумме полей магнитного и электрического диполей.*

### Введение

Теоретическое исследование композитных материалов является одним из приоритетных направлений фундаментальной науки. Композиты отличаются большим разнообразием в силу разнообразия структурного и химического состава материалов, которые определяют электродинамические свойства выполненных из них технических устройств [1]. В последнее время большое внимание в научной литературе уделяется исследованию биизотропных и киральных метаматериалов. В работах [2–5] приведены обзоры публикаций по данной тематике. Важными для приложений техническими устройствами являются плоские экраны, выполненные из биизотропных материалов. Такие системы служат для экранирования электромагнитных полей, а также для преобразования структуры падающего поля при прохождении через экран. В ряде работ аналитически и численно исследованы экранирующие свойства биизотропных экранов при воздействии на них монохроматических полей различных типов: плоских полей [6], полей источников [7] и пучков волн [8]. Начиная с пионерских работ В.Г. Веселаго [9, 10], проводятся исследования плоских пластин, выполненных из метаматериалов. Такие структуры обладают уникальными свойствами фокусировки и, по сути, являются фокусирующими линзами. Электромагнитное поле источника, расположенного по одну сторону пластины из метаматериала, локализуется по другую сторону пластины в виде фокуса. Исследовались пластины из бианизотропного материала [11]. Разработана детальная лучевая теория фокусирующих однослойных и двухслойных линз [5, 12], в которой анализируются законы отражения и преломления полей на плоскостях раздела сред из вакуума и метаматериала. Разрабатывались другие подходы к моделированию таких структур [5]. Исследованы аналитически [13] и численно [7, 8] фокусирующие свойства экранов из биизотропных материалов.

В настоящей статье получено точное аналитическое решение краевой задачи проникновения электромагнитного поля магнитного диполя через плоскую бесконечную линзу из метаматериала. Применен метод с использованием двухсторонних граничных условий [14] (нелокальных граничных условий) на плоском биизотропном экране и метод теорем сложения, связывающих базисные сферические и цилиндрические электромагнитные поля. Описано многообразие материальных параметров линзы, для которой поле магнитного диполя, расположенного по одну сторону слоя, фокусируется в поле магнитного диполя по другую сторону слоя. Детально исследовано отраженное поле в полупространстве, в котором находится точечный источник поля. Показано, что при определенных условиях в этом полупространстве образуется второй фокус.

### 1. Постановка задачи

В пространстве  $R^3$  с декартовой системой координат  $Oxuz$  размещен плоский экран  $D(0 < z < \Delta)$ , который ограничен плоскостями  $\Gamma_1(z=0)$ ,  $\Gamma_2(z=\Delta)$ . Слой выполнен из биизотропного материала, характеризуемого материальными комплексными параметрами

$\varepsilon = \varepsilon_r \varepsilon_0$ ,  $\mu = \mu_r \mu_0$ ,  $G = G_r/c$ ,  $Z = Z_r/c$ , где  $\varepsilon_0, \mu_0$  – электрическая и магнитная постоянные,  $c$  – скорость света в вакууме. В полупространстве  $D_1(z < \Delta)$  на оси  $Oz$  в точке  $O_1(0, 0, -h)$ ,  $h > 0$ , размещен магнитный диполь, колеблющийся с круговой частотой  $\omega = 2\pi f$  и возбуждающий первичное поле  $\vec{E}_0, \vec{H}_0$ . В результате взаимодействия поля диполя с экраном  $D$  образуются поля:  $\vec{E}'_1, \vec{H}'_1$  – поле в  $D_1$ , отраженное от экрана  $D$ ;  $\vec{E}_1 = \vec{E}_0 + \vec{E}'_1$ ,  $\vec{H}_1 = \vec{H}_0 + \vec{H}'_1$  – суммарное поле в области  $D_1$ ;  $\vec{E}, \vec{H}$  – поле в слое  $D$ ;  $\vec{E}_2, \vec{H}_2$  – поле, проникшее в область  $D_2(z > \Delta)$ . Рассматривается временная зависимость полей вида  $\exp(-i\omega t)$ .

Сформулируем краевую задачу, моделирующую проникновение первичного поля  $\vec{E}_0, \vec{H}_0$  через экран  $D$  [13].

*Краевая задача 1.* При заданном поле  $\vec{E}_0, \vec{H}_0$  требуется определить поля  $\vec{E}'_1, \vec{H}'_1$ ;  $\vec{E}_2, \vec{H}_2$ ;  $\vec{E}, \vec{H}$ , которые удовлетворяют:

– уравнениям

$$\text{rot } \vec{E}_j = i\omega\mu_0 \vec{H}_j, \quad \text{rot } \vec{H}_j = -i\omega\varepsilon_0 \vec{E}_j \quad \text{в } D_j, \quad (1)$$

$$\text{rot } \vec{E} = i\omega(\mu \vec{H} + Z \vec{E}), \quad \text{rot } \vec{H} = -i\omega(\varepsilon \vec{E} + G \vec{H}) \quad \text{в } D; \quad (2)$$

– граничным условиям непрерывности тангенциальных составляющих полей на плоскостях  $\Gamma_1, \Gamma_2$

$$\left. (\vec{E}_{1r} - \vec{E}_r) \right|_{z=0} = 0, \quad \left. (\vec{H}_{1r} - \vec{H}_r) \right|_{z=0} = 0, \quad \left. (\vec{E}_{2r} - \vec{E}) \right|_{z=\Delta} = 0, \quad \left. (\vec{H}_{2r} - \vec{H}_r) \right|_{z=\Delta} = 0; \quad (3)$$

– условиям излучения на бесконечности

$$\lim_{r \rightarrow \infty} r \left( \frac{\partial \vec{E}'_1}{\partial r} - ik_0 \vec{E}'_1 \right) = 0, \quad z < 0, \quad \lim_{r \rightarrow \infty} r \left( \frac{\partial \vec{E}_2}{\partial r} - ik_0 \vec{E}_2 \right) = 0, \quad z > \Delta, \quad (4)$$

где  $k_0 = \frac{\omega}{c}$ .  $r$  – радиальная координата сферической системы координат  $Or\theta\varphi$ .

## 2. Представление решения задачи через цилиндрические базисные поля

Рассмотрим магнитный и электрический диполи, расположенные в начале координат  $O$  и ориентированные вдоль оси  $Oz$ :

– магнитный диполь ( $TE$ -диполь, электрический вектор параллелен экрану  $D$ ) [15, с. 174]

$$\vec{E}_m(O) = A_0 \tilde{m}_{01}(\vec{r}, k_0), \quad \vec{H}_m(O) = h_0 A_0 \tilde{n}_{01}(\vec{r}, k_0), \quad (5)$$

где  $\vec{r} = (r, \theta, \varphi)$ ,  $h_0 = \frac{1}{iZ_0}$ ,  $Z_0 = \sqrt{\frac{\mu_0}{\varepsilon_0}}$ ;

– электрический диполь ( $TH$ -диполь, магнитный вектор параллелен экрану  $D$ ) [15, с. 166]

$$\vec{E}_e(O) = B_0 \tilde{n}_{01}(\vec{r}, k_0), \quad \vec{H}_e(O) = h_0 B_0 \tilde{m}_{01}(\vec{r}, k_0), \quad (6)$$

где  $A_0, B_0$  – амплитуды диполей;  $\tilde{m}_{01}, \tilde{n}_{01}$  – сферические поля [15, с.118].

Учитывая осевую симметрию полей диполей, представим поля (5), (6) через базисные цилиндрические поля [15, с. 130]:

$$\vec{M}_m^{(\mp 1)}(\vec{\rho}, \lambda, k_0) = \vec{V}_m^{(1)}(\lambda \rho) e^{\mp v_0(\lambda)z} \Phi_m, \quad (7)$$

$$\vec{M}_m^{(\mp 2)}(\vec{\rho}, \lambda, k_0) = \frac{1}{k_0} \left( \mp v_0(\lambda) \vec{V}_m^{(2)}(\lambda \rho) + \lambda J_m(\lambda \rho) \vec{e}_z \right) e^{\mp v_0(\lambda)z} \Phi_m,$$

где  $\vec{V}_m^{(1)}(\lambda \rho) = \frac{im}{\lambda \rho} J_m(\lambda \rho) \vec{e}_\rho - J'_m(\lambda \rho) \vec{e}_\varphi$ ,  $\vec{V}_m^{(2)}(\lambda \rho) = J'_m(\lambda \rho) \vec{e}_\rho + \frac{im}{\lambda \rho} J_m(\lambda \rho) \vec{e}_\varphi$ ,  $v_0(\lambda) = \sqrt{\lambda^2 - k_0^2}$  при  $\lambda \geq k_0$ ;  $v_0(\lambda) = -i\sqrt{k_0^2 - \lambda^2}$  при  $k_0 > \lambda$ ,  $0 \leq \lambda < \infty$ ;  $\vec{e}_\rho, \vec{e}_\varphi, \vec{e}_z$  – орты цилиндрической системы координат,  $J_m(\cdot)$  – функции Бесселя,  $J'_m(\cdot)$  – производная,  $\vec{\rho} = (\rho, z, \varphi)$ ,  $\Phi_m = e^{im\varphi}$ .

Используя теоремы сложения [15, с.153], представим поля диполей (5), (6) в интегральном виде через поля (7) при  $m = 0$ :

$$\begin{aligned} \vec{E}_m(O) &= \begin{cases} A_0 \int_0^\infty C(\lambda) \vec{M}_0^{(-1)}(\vec{\rho}, \lambda, k_0) d\lambda, & z > 0, \\ A_0 \int_0^\infty C(\lambda) \vec{M}_0^{(+1)}(\vec{\rho}, \lambda, k_0) d\lambda, & z < 0, \end{cases} \\ \vec{H}_m(O) &= \begin{cases} h_0 A_0 \int_0^\infty C(\lambda) \vec{M}_0^{(-2)}(\vec{\rho}, \lambda, k_0) d\lambda, & z > 0, \\ h_0 A_0 \int_0^\infty C(\lambda) \vec{M}_0^{(+2)}(\vec{\rho}, \lambda, k_0) d\lambda, & z < 0; \end{cases} \\ \vec{E}_s(O) &= \begin{cases} B_0 \int_0^\infty C(\lambda) \vec{M}_0^{(-2)}(\vec{\rho}, \lambda, k_0) d\lambda, & z > 0, \\ B_0 \int_0^\infty C(\lambda) \vec{M}_0^{(+2)}(\vec{\rho}, \lambda, k_0) d\lambda, & z < 0, \end{cases} \\ \vec{H}_s(O) &= \begin{cases} h_0 B_0 \int_0^\infty C(\lambda) \vec{M}_0^{(-1)}(\vec{\rho}, \lambda, k_0) d\lambda, & z > 0, \\ h_0 B_0 \int_0^\infty C(\lambda) \vec{M}_0^{(+1)}(\vec{\rho}, \lambda, k_0) d\lambda, & z < 0, \end{cases} \end{aligned} \quad (8)$$

$$\text{где } C(\lambda) = -\frac{i\lambda^2}{k_0^2 v_0(\lambda)}.$$

Для диполей (5), (6), расположенных в произвольной точке  $O_s(0, 0, h_s)$ ,  $s = 1, 2, 3$ , на оси  $Oz$ , учитывая сдвиг системы координат  $z = z_s + h_s$  в формулах (8), получим интегральные представления

$$\vec{E}_m(O_s) = \begin{cases} \int_0^\infty A(\lambda, h_s) \vec{M}_0^{(-1)}(\vec{\rho}, \lambda, k_0) d\lambda, & z > h_s, \\ \int_0^\infty A(\lambda, -h_s) \vec{M}_0^{(+1)}(\vec{\rho}, \lambda, k_0) d\lambda, & z < h_s, \end{cases}$$

$$\vec{\mathbf{H}}_m(O_s) = \begin{cases} h_0 \int_0^{\infty} A(\lambda, h_s) \vec{M}_0^{(-2)}(\vec{\rho}, \lambda, k_0) d\lambda, & z > h_s, \\ h_0 \int_0^{\infty} A(\lambda, -h_s) \vec{M}_0^{(+2)}(\vec{\rho}, \lambda, k_0) d\lambda, & z < h_s; \end{cases}$$

$$\vec{\mathbf{E}}_y(O_s) = \begin{cases} \int_0^{\infty} B(\lambda, h_s) \vec{M}_0^{(-2)}(\vec{\rho}, \lambda, k_0) d\lambda, & z > h_s, \\ \int_0^{\infty} B(\lambda, -h_s) \vec{M}_0^{(+2)}(\vec{\rho}, \lambda, k_0) d\lambda, & z < h_s, \end{cases} \quad (9)$$

$$\vec{\mathbf{H}}_y(O_s) = \begin{cases} h_0 \int_0^{\infty} B(\lambda, h_s) \vec{M}_0^{(-1)}(\vec{\rho}, \lambda, k_0) d\lambda, & z > h_s, \\ h_0 \int_0^{\infty} B(\lambda, -h_s) \vec{M}_0^{(+1)}(\vec{\rho}, \lambda, k_0) d\lambda, & z < h_s, \end{cases}$$

где  $A(\lambda, h_s) = A_0 C(\lambda) e^{v_0(\lambda)h_s}$ ,  $B(\lambda, h_s) = B_0 C(\lambda) e^{v_0(\lambda)h_s}$ . (10)

Для поля магнитного диполя, расположенного в точке  $O_1(0, 0, -h) \in D_1$ , из формул (9) следует интегральное представление

$$\vec{\mathbf{E}}_0 = \vec{\mathbf{E}}_m(O_1) = \int_0^{\infty} A(\lambda, -h) \vec{M}_0^{(-1)}(\vec{\rho}, \lambda, k_0) d\lambda, \quad z > -h,$$

$$\vec{\mathbf{H}}_0 = \vec{\mathbf{H}}_m(O_1) = h_0 \int_0^{\infty} A(\lambda, -h) \vec{M}_0^{(-2)}(\vec{\rho}, \lambda, k_0) d\lambda. \quad (11)$$

Учитывая структуру (11) первичного поля, решение задачи (1)–(4) также представим в интегральном виде:

$$\vec{\mathbf{E}}_1' = \int_0^{\infty} \vec{E}_1'(\lambda) d\lambda, \quad \vec{\mathbf{H}}_1' = \int_0^{\infty} \vec{H}_1'(\lambda) d\lambda, \quad z < 0;$$

$$\vec{\mathbf{E}}_2 = \int_0^{\infty} \vec{E}_2(\lambda) d\lambda, \quad \vec{\mathbf{H}}_2 = \int_0^{\infty} \vec{H}_2(\lambda) d\lambda, \quad z > \Delta;$$

$$\vec{\mathbf{E}} = \int_0^{\infty} \vec{E}(\lambda) d\lambda, \quad \vec{\mathbf{H}} = \int_0^{\infty} \vec{H}(\lambda) d\lambda, \quad 0 < z < \Delta; \quad (12)$$

$$\vec{\mathbf{E}}_0 = \int_0^{\infty} \vec{E}_0(\lambda) d\lambda, \quad \vec{\mathbf{H}}_0 = \int_0^{\infty} \vec{H}_0(\lambda) d\lambda, \quad z > -h;$$

$$\vec{E}_1 = \vec{E}_1' + \vec{E}_0, \quad \vec{H}_1 = \vec{H}_1' + \vec{H}_0,$$

где

$$\begin{aligned} \vec{E}_1'(\lambda) &= x_1(\lambda) \vec{M}_0^{(+1)} + y_1(\lambda) \vec{M}_0^{(+2)}, & \vec{H}_1'(\lambda) &= h_0 (x_1(\lambda) \vec{M}_0^{(+2)} + y_1(\lambda) \vec{M}_0^{(+1)}), \\ \vec{E}_2(\lambda) &= x_2(\lambda) \vec{M}_0^{(-1)} + y_2(\lambda) \vec{M}_0^{(-2)}, & \vec{H}_2(\lambda) &= h_0 (x_2(\lambda) \vec{M}_0^{(-2)} + y_2(\lambda) \vec{M}_0^{(-1)}), \\ \vec{E}_0(\lambda) &= a(\lambda) \vec{M}_0^{(-1)} + b(\lambda) \vec{M}_0^{(-2)}, & \vec{H}_0(\lambda) &= h_0 (a(\lambda) \vec{M}_0^{(-2)} + b(\lambda) \vec{M}_0^{(-1)}), \\ a(\lambda) &= A(\lambda, -h), & b(\lambda) &= 0. \end{aligned} \quad (13)$$

Формулы для полей  $\vec{E}(\lambda), \vec{H}(\lambda)$  в области  $D$  опускаем, так как основная задача состоит в определении полей до экрана и за экраном в областях  $D_1$  и  $D_2$ .

Поля  $\vec{E}_1, \vec{H}_1; \vec{E}_2, \vec{H}_2; \vec{E}, \vec{H}$  удовлетворяют условиям (1)–(3) краевой задачи 1. Условия (4) выполнены для полей (12), так как электромагнитная энергия поля  $\vec{E}'_1, \vec{H}'_1$  (13) распространяется в отрицательном направлении оси  $Oz$ , а поля  $\vec{E}_2, \vec{H}_2$  (13) – в положительном направлении оси  $Oz$ .

### 3. Аналитическое решение краевой задачи экранирования

Для аналитического вычисления амплитуд  $x_j(\lambda), y_j(\lambda)$  полей (13) воспользуемся двухсторонними граничными условиями [14], эквивалентными граничным условиям (3). Применение двухсторонних граничных условий позволяет исключить из рассмотрения поля  $\vec{E}, \vec{H}$  в слое  $D$ . Для формулировки двухсторонних условий на экране, связывающих поля  $\vec{E}_1, \vec{H}_1$  и  $\vec{E}_2, \vec{H}_2$  по обе стороны экрана, запишем тангенциальные составляющие полей на плоскостях  $\Gamma_1$  и  $\Gamma_2$  в базисе  $\vec{V}_0^{(1)}, \vec{V}_0^{(2)}$  (7):

$$\begin{aligned}\vec{E}_{1\tau}(\lambda)\Big|_{z=0} &= E_{1V_1}(\lambda)\vec{V}_0^{(1)}(\lambda\rho) + E_{1V_2}(\lambda)\vec{V}_0^{(2)}(\lambda\rho), \\ \vec{H}_{1\tau}(\lambda)\Big|_{z=0} &= H_{1V_1}(\lambda)\vec{V}_0^{(1)}(\lambda\rho) + H_{1V_2}(\lambda)\vec{V}_0^{(2)}(\lambda\rho); \\ \vec{E}_{2\tau}(\lambda)\Big|_{z=\Delta} &= E_{2V_1}(\lambda)\vec{V}_0^{(1)}(\lambda\rho) + E_{2V_2}(\lambda)\vec{V}_0^{(2)}(\lambda\rho), \\ \vec{H}_{2\tau}(\lambda)\Big|_{z=\Delta} &= H_{2V_1}(\lambda)\vec{V}_0^{(1)}(\lambda\rho) + H_{2V_2}(\lambda)\vec{V}_0^{(2)}(\lambda\rho),\end{aligned}$$

где

$$\begin{aligned}E_{1V_1} &= E'_{1V_1} + E_{0V_1} = x_1(\lambda) + a(\lambda), \quad E_{1V_2} = E'_{1V_2} + E_{0V_2} = \bar{v}_0(\lambda)(y_1(\lambda) - b(\lambda)), \\ H_{1V_1} &= H'_{1V_1} + H_{0V_1} = h_0(y_1(\lambda) + b(\lambda)), \quad H_{1V_2} = H'_{1V_2} + H_{0V_2} = h_0\bar{v}_0(\lambda)(x_1(\lambda) - a(\lambda)); \\ E_{2V_1} &= x_2(\lambda)F(\lambda), \quad E_{2V_2} = -\bar{v}_0(\lambda)y_2(\lambda)F(\lambda), \\ H_{2V_1} &= h_0y_2(\lambda)F(\lambda), \quad H_{2V_2} = -h_0\bar{v}_0(\lambda)x_2(\lambda)F(\lambda), \\ F(\lambda) &= e^{-v_0(\lambda)\Delta}, \quad \bar{v}_0(\lambda) = \frac{v_0(\lambda)}{k_0}.\end{aligned}\tag{14}$$

*Краевая задача 2.* При заданном поле  $\vec{E}_0, \vec{H}_0$  (13) требуется определить поля  $\vec{E}'_1, \vec{H}'_1; \vec{E}_2, \vec{H}_2$ , которые удовлетворяют:

– уравнениям

$$\text{rot } \vec{E}_j = i\omega\mu_0\vec{H}_j, \quad \text{rot } \vec{H}_j = -i\omega\varepsilon_0\vec{E}_j \quad \text{в } D_j;$$

– двухстороннему граничному условию [14]

$$\vec{U}_1\Big|_{z=0} = \hat{C}(\lambda)\vec{U}_2\Big|_{z=\Delta};\tag{15}$$

– условиям излучения в областях  $D_j$ .

Здесь

$$\hat{C} = \begin{pmatrix} C_{11} & C_{12} & C_{13} & C_{14} \\ C_{21} & C_{22} & C_{23} & C_{24} \\ C_{31} & C_{32} & C_{33} & C_{34} \\ C_{41} & C_{42} & C_{43} & C_{44} \end{pmatrix}, \quad \vec{U}_j = \begin{pmatrix} E_{j\nu_1} \\ H_{j\nu_2} \\ E_{j\nu_2} \\ H_{j\nu_1} \end{pmatrix}, \quad j=1, 2; \quad (16)$$

$$\begin{aligned} C_{11} &= p(p_1 C_2 - p_2 C_1), & C_{12} &= p\left(q_1 \frac{k_1}{v_1} S_1 - q_2 \frac{k_2}{v_2} S_2\right), \\ C_{13} &= p\left(q_2 \frac{p_1 k_2}{v_2} S_2 - q_1 \frac{p_2 k_1}{v_1} S_1\right), & C_{14} &= p(C_1 - C_2), \\ C_{21} &= pp_1 p_2 \left(q_2 \frac{v_2}{k_2} S_2 - q_1 \frac{v_1}{k_1} S_1\right), & C_{22} &= p(p_1 C_1 - p_2 C_2), \\ C_{23} &= pp_1 p_2 (C_2 - C_1), & C_{24} &= p\left(\frac{p_1 v_1}{q_1 k_1} S_1 - \frac{p_2 v_2}{q_2 k_2} S_2\right), \\ C_{31} &= p\left(\frac{p_1 v_2}{q_2 k_2} S_2 - \frac{p_2 v_1}{q_1 k_1} S_1\right), & C_{32} &= p(C_1 - C_2), \\ C_{33} &= p(p_1 C_2 - p_2 C_1), & C_{34} &= p\left(\frac{v_1}{q_1 k_1} S_1 - \frac{v_2}{q_2 k_2} S_2\right), \\ C_{41} &= pp_1 p_2 (C_2 - C_1), & C_{42} &= p\left(q_1 p_1 \frac{k_1}{v_1} S_1 - q_2 p_2 \frac{k_2}{v_2} S_2\right), \\ C_{43} &= pp_1 p_2 \left(q_2 \frac{k_2}{v_2} S_2 - q_1 \frac{k_1}{v_1} S_1\right), & C_{44} &= p(p_1 C_1 - p_2 C_2), \end{aligned} \quad (17)$$

$$C_j = \text{ch}(\xi_j), \quad S_j = \text{sh}(\xi_j), \quad \xi_j = v_j(\lambda)\Delta, \quad v_j(\lambda) = \sqrt{\lambda^2 - k_j^2}, \quad -\frac{\pi}{2} \leq \arg v_j < \frac{\pi}{2};$$

$$k_j = k_0 \bar{k}_j, \quad f_0 = k_0 \bar{f}_0, \quad f_j = k_0 \bar{f}_j, \quad g = k_0^2 \bar{g}, \quad g_j = k_0 \bar{g}_j, \quad a = k_0 \bar{a}, \quad b = k_0 \bar{b}, \quad p_j = \frac{\bar{p}_j}{Z_0},$$

$$q_j = \frac{\bar{g}}{k_j \bar{g}_j}, \quad v_j = k_0 \bar{v}_j, \quad p = \frac{1}{p_1 - p_2} = Z_0 \bar{p}, \quad k_0 = \frac{\omega}{c},$$

где безразмерные величины

$$\begin{aligned} \bar{k}_j &= \sqrt{\bar{g} + \frac{1}{2}\bar{a}^2 + \bar{a}\bar{f}_j}, \quad \bar{f}_j = (-1)^j \bar{f}_0, \quad \bar{f}_0 = \sqrt{\varepsilon_r \mu_r - \bar{b}^2}, \\ \bar{g} &= \varepsilon_r \mu_r - G_r Z_r, \quad \bar{g}_j = \bar{f}_j - \frac{1}{2}\bar{a}, \quad \bar{p}_j = \frac{1}{\mu_r} \left( \frac{i\bar{g}}{\bar{g}_j} - Z_r \right), \quad \bar{a} = i(G_r - Z_r), \\ \bar{b} &= \frac{1}{2}(G_r + Z_r), \quad \bar{v}_j = \sqrt{\bar{\lambda}^2 - \bar{k}_j^2}, \quad \bar{\lambda} = \frac{\lambda}{k_0}, \quad \bar{p} = \frac{1}{\bar{p}_1 - \bar{p}_2}. \quad \blacksquare \end{aligned} \quad (18)$$

Для решения краевой задачи подставим компоненты (14) в условие (15), получим систему алгебраических уравнений для определения амплитуд  $x_j(\lambda)$ ,  $y_j(\lambda)$ . Решая систему аналитически, находим [7]

$$x_2(\lambda) = \frac{2G}{d}(aQ_{22} - bQ_{12}), \quad y_2(\lambda) = \frac{2G}{d}(bQ_{11} - aQ_{21}),$$

$$x_1(\lambda) = \frac{2}{d} \left[ (C_{11} - h_0 \bar{v}_0 C_{12})(aQ_{22} - bQ_{12}) + (\bar{v}_0 C_{13} - h_0 C_{14})(aQ_{21} - bQ_{11}) \right] - a, \quad (19)$$

$$y_1(\lambda) = \frac{2}{d} \left[ \left( \frac{1}{h_0} C_{41} - \bar{v}_0 C_{42} \right) (aQ_{22} - bQ_{12}) + \left( \frac{\bar{v}_0}{h_0} C_{43} - C_{44} \right) (aQ_{21} - bQ_{11}) \right] - b;$$

$$Q_{11} = C_{11} - h_0 \bar{v}_0 C_{12} - \frac{1}{h_0 \bar{v}_0} C_{21} + C_{22},$$

$$Q_{12} = -\bar{v}_0 C_{13} + h_0 C_{14} + \frac{1}{h_0} C_{23} - \frac{1}{\bar{v}_0} C_{24},$$

$$Q_{21} = -\frac{1}{\bar{v}_0} C_{31} + h_0 C_{32} + \frac{1}{h_0} C_{41} - \bar{v}_0 C_{42}, \quad (20)$$

$$Q_{22} = C_{33} - \frac{h_0}{\bar{v}_0} C_{34} - \frac{\bar{v}_0}{h_0} C_{43} + C_{44}, \quad d = Q_{11}Q_{22} - Q_{12}Q_{21},$$

$$G = \exp(v_0 \Delta), \quad \bar{v}_0 = \frac{v_0}{k_0}.$$

Амплитуды (19) полей (13) определены. Подставляя поля (13) в интегралы (12), получим решение краевой задачи 1:

$$\vec{\mathbf{E}}_2 = \int_0^\infty A(\lambda, -h) \frac{2G}{d} \left( Q_{22} \vec{M}_0^{(-1)}(\vec{\rho}, \lambda, k_0) - Q_{21} \vec{M}_0^{(-2)}(\vec{\rho}, \lambda, k_0) \right) d\lambda, \quad (21)$$

$$\vec{\mathbf{H}}_2 = h_0 \int_0^\infty A(\lambda, -h) \frac{2G}{d} \left( Q_{22} \vec{M}_0^{(-2)}(\vec{\rho}, \lambda, k_0) - Q_{21} \vec{M}_0^{(-1)}(\vec{\rho}, \lambda, k_0) \right) d\lambda;$$

$$\vec{\mathbf{E}}'_1 = \int_0^\infty A(\lambda, -h) \frac{2}{d} \left( P_{22} \vec{M}_0^{(+1)}(\vec{\rho}, \lambda, k_0) + P_{21} \vec{M}_0^{(+2)}(\vec{\rho}, \lambda, k_0) \right) d\lambda, \quad (22)$$

$$\vec{\mathbf{H}}'_1 = h_0 \int_0^\infty A(\lambda, -h) \frac{2}{d} \left( P_{22} \vec{M}_0^{(+2)}(\vec{\rho}, \lambda, k_0) + P_{21} \vec{M}_0^{(+1)}(\vec{\rho}, \lambda, k_0) \right) d\lambda,$$

где  $P_{22} = (C_{11} - h_0 \bar{v}_0 C_{12})Q_{22} + (\bar{v}_0 C_{13} - h_0 C_{14})Q_{21} - \frac{1}{2}d$ ,  $P_{21} = \left( \frac{1}{h_0} C_{41} - \bar{v}_0 C_{42} \right) Q_{22} + \left( \frac{\bar{v}_0}{h_0} C_{43} - C_{44} \right) Q_{21}$ .

В случае поля магнитного диполя (11) в формулах (19)  $a = A(\lambda, -h)$ ,  $b = 0$ .

#### 4. Фокусировка электромагнитного поля за экраном

Для получения фокусирующих свойств на материальные параметры экрана  $D$  наложим ряд условий [13]:

– первое условие

$$G_r = Z_r = \chi, \quad (23)$$

где  $\chi$  – комплексное число;

– второе условие

$$\bar{g} = \varepsilon_r \mu_r - G_r Z_r = \varepsilon_r \mu_r - \chi^2 = 1. \quad (24)$$

Применение условий (23), (24) упрощает величины (18):

$$\bar{a} = 0, \quad \bar{g} = 1, \quad \bar{f}_0 = 1, \quad \bar{k}_1 = \bar{k}_2 = 1, \quad g_1 = -1, \quad g_2 = 1,$$

$$q_1 = -1, \quad q_2 = 1, \quad \bar{p}_1 = -\frac{i}{\mu_r}(1 - i\chi), \quad \bar{p}_2 = \frac{i}{\mu_r}(1 + i\chi),$$

$$\bar{p} = \frac{i\mu_r}{2}, \quad \bar{f}_1 = -1, \quad \bar{f}_2 = 1, \quad \bar{b} = \chi, \quad \bar{v}_1 = \bar{v}_2 = \bar{v}_0.$$

Для матричных элементов (17) получим значения

$$C_{11} = C_{22} = C_{33} = C_{44} = \text{ch}(\xi_0), \quad C_{14} = C_{23} = C_{32} = C_{41} = 0,$$

$$C_{13} = -\frac{i\chi}{\bar{v}_0} \text{sh}(\xi_0), \quad C_{12} = -\frac{\mu_r}{\bar{v}_0 h_0} \text{sh}(\xi_0), \quad C_{21} = -h_0 \varepsilon_r \bar{v}_0 \text{sh}(\xi_0), \quad (25)$$

$$C_{24} = i\chi \bar{v}_0 \text{sh}(\xi_0), \quad C_{31} = -i\chi \bar{v}_0 \text{sh}(\xi_0), \quad C_{34} = -\frac{\mu_r \bar{v}_0}{h_0} \text{sh}(\xi_0),$$

$$C_{42} = \frac{i\chi}{\bar{v}_0} \text{sh}(\xi_0), \quad C_{43} = -\frac{h_0 \varepsilon_r}{\bar{v}_0} \text{sh}(\xi_0), \quad \xi_0 = v_0 \Delta.$$

Коэффициенты (20) упрощаются:

$$Q_{11} = Q_{22} = Q = 2 \text{ch}(\xi_0) + (\varepsilon_r + \mu_r) \text{sh}(\xi_0), \quad (26)$$

$$Q_{12} = Q_{21} = 0, \quad d = Q^2, \quad G = e^{i\xi_0}.$$

Введем параметр фокусировки:

$$\tau = 1 + \frac{1}{2}(\varepsilon_r + \mu_r), \quad (27)$$

тогда

$$Q = e^{-\xi_0} \left( 2 - \tau \left( 1 - e^{2\xi_0} \right) \right).$$

Далее вычислим электрическое поле (21) в области  $D_2$  с учетом формул (10), (26):

$$\begin{aligned}\vec{E}_2 &= \int_0^\infty A(\lambda, -h) \frac{2G}{Q} \vec{M}_0^{(-1)}(\vec{\rho}, \lambda, k_0) d\lambda = \\ &= 2 \int_0^\infty A(\lambda, h_f) \frac{\exp\left(\nu_0(2\Delta - h - h_f)\right)}{2 - \tau(1 - e^{2\xi_0})} \vec{M}_0^{(-1)}(\vec{\rho}, \lambda, k_0) d\lambda,\end{aligned}$$

где  $h_f$  – произвольная величина.

Введем еще одно (третье) условие фокусировки:

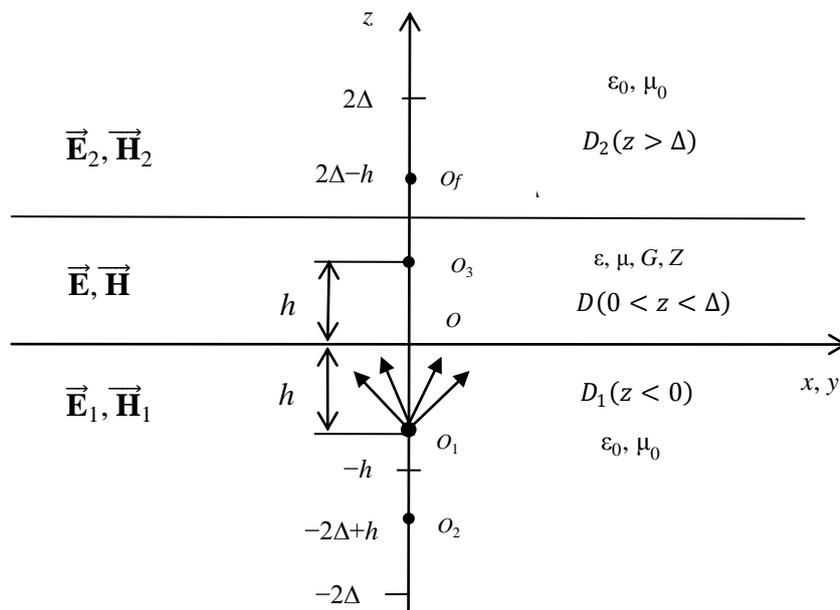
$$\tau = 1 + \frac{1}{2}(\varepsilon_r + \mu_r) = 0. \quad (28)$$

Величину  $h_f$  выберем из условия  $h_f = 2\Delta - h$ , тогда с учетом (9) получим

$$\vec{E}_2 = \int_0^\infty A(\lambda, h_f) \vec{M}_0^{(-1)}(\vec{\rho}, \lambda, k_0) d\lambda = \vec{E}_m(O_f), \quad (29)$$

где  $O_f$  – точка с координатами  $(0, 0, h_f)$ .

**Выводы.** При выполнении условий (23), (24), (28) электромагнитное поле за экраном в области  $D_2$  является полем магнитного диполя, расположенного в точке  $O_f$  (рисунок). При ограничении  $0 < h < \Delta$  точка  $O_f$  расположена в области  $D_2$  на расстоянии  $\Delta - h$  от экрана  $D$ , т. е. поле  $\vec{E}_2, \vec{H}_2$  фокусируется в точке  $O_f$ , которая является фокусом первичного поля  $\vec{E}_0, \vec{H}_0$ . В случае  $h > \Delta$  точка  $O_f$  является фиктивным фокусом, лежащим вне области  $D_2$ . ■



Плоская линза с точками фокусировки поля:

$O_1$  – источник поля;  $O_f, O_2$  – фокусы;  $O_3$  – фиктивный фокус

Таким образом, условия (23), (24), (28) выделяют из параметров биизотропного материала фокусирующие параметры. Для описания параметров фокусировки выберем произвольный комплексный параметр  $\delta$  и определим два комплексных многообразия материальных параметров экрана  $D$ :

$$\begin{aligned}\varepsilon_r = \delta, \quad \mu_r = -2 - \delta, \quad G_r = Z_r = i(1 + \delta); \\ \varepsilon_r = \delta, \quad \mu_r = -2 - \delta, \quad G_r = Z_r = -i(1 + \delta),\end{aligned}\tag{30}$$

где параметр  $\delta$  параметризует многообразия. При  $\delta = -1$  получим метаматериал Веселаго:  $\varepsilon_r = -1$ ,  $\mu_r = -1$ ,  $G_r = Z_r = 0$ .

### 5. Фокусировка электромагнитного поля перед экраном

Вычислим отраженное поле (22) в области  $D_1$ . Учитывая равенства (25), (26), найдем величины

$$\begin{aligned}P_{22} = Q \left( C_{11} - h_0 \bar{\nu} C_{12} - \frac{1}{2} Q \right) = \frac{Q}{2} (\mu_r - \varepsilon_r) \operatorname{sh}(\xi_0) = \frac{Q}{4} (\mu_r - \varepsilon_r) (e^{\xi_0} - e^{-\xi_0}), \\ P_{21} = -\nu_0 Q C_{42} = -i\chi Q \operatorname{sh}(\xi_0) = -\frac{i\chi}{2} Q (e^{\xi_0} - e^{-\xi_0}).\end{aligned}$$

В результате получим

$$\begin{aligned}\vec{\mathbf{E}}'_1 = \int_0^\infty \frac{2}{Q} A(\lambda, -h) (e^{\xi_0} - e^{-\xi_0}) (\alpha_1 \vec{M}_0^{(+1)} - \alpha_2 \vec{M}_0^{(+2)}) d\lambda = \\ = \int_0^\infty \frac{2A(\lambda, -h) (e^{2\nu_0 \Lambda} - 1)}{2 - \tau(1 - e^{2\xi_0})} (\alpha_1 M_0^{(+1)} - \alpha_2 \vec{M}_0^{(+2)}) d\lambda,\end{aligned}$$

где  $\alpha_1 = \frac{1}{4}(\mu_r - \varepsilon_r)$ ,  $\alpha_2 = \frac{i\chi}{2}$ .

При фокусировке (28) с учетом формулы (10)

$$\vec{\mathbf{E}}'_1 = \int_0^\infty A(\lambda, h_f) (\alpha_1 \vec{M}_0^{(+1)} - \alpha_2 \vec{M}_0^{(+2)}) d\lambda - \int_0^\infty A(\lambda, -h) (\alpha_1 M_0^{(+1)} - \alpha_2 \vec{M}_0^{(+2)}) d\lambda.\tag{31}$$

Рассмотрим точки  $O_2(0, 0, -h_f)$ ,  $O_3(0, 0, h)$ . Полагая  $h_2 = -h_f$ ,  $h_3 = h$  и сравнивая формулу (31) с интегральными представлениями магнитных и электрических диполей (9) в нижнем полупространстве, получим выражение поля (31) через сумму полей диполей:

$$\begin{aligned}\vec{\mathbf{E}}'_1 = \alpha_1 \vec{\mathbf{E}}_m(O_2) - \alpha_2 \vec{\mathbf{E}}_e(O_2) - \alpha_1 \vec{\mathbf{E}}_m(O_3) + \alpha_2 \vec{\mathbf{E}}_e(O_3) = \\ = A_0 (\alpha_1 \tilde{m}_{01}(\vec{r}_2, k_0) - \alpha_2 \tilde{n}_{01}(\vec{r}_2, k_0) - \alpha_1 \tilde{m}_{01}(\vec{r}_3, k_0) + \alpha_2 \tilde{n}_{01}(\vec{r}_3, k_0)).\end{aligned}$$

Отметим, что для метаматериала ( $\delta = -1$ ) коэффициенты  $\alpha_1 = \alpha_2 = 0$ , т. е. отраженное поле  $\vec{\mathbf{E}}'_1 = 0$ ,  $\vec{\mathbf{H}}'_1 = 0$  в  $D_1$  и в области  $D_1$  присутствует только первичное поле  $\vec{\mathbf{E}}_0, \vec{\mathbf{H}}_0$  с источником в точке  $O_1$ .

**Выводы.** При выполнении условий (23), (24), (28) для материальных параметров экрана (30) отраженное электромагнитное поле перед экраном в области  $D_1$  является суммой полей двух магнитных и двух электрических диполей, попарно расположенных в точках  $O_2$  и  $O_3$  (см. рис. 1). При условии  $0 < h < 2\Delta$  точка  $O_2$  расположена в области  $D_1$  на расстоянии  $h_f$  от экрана  $D$ , т. е. поле  $\vec{\mathbf{E}}'_1, \vec{\mathbf{H}}'_1$  фокусируется в точке  $O_2$  в виде суммы магнитного и электричес-

кого диполей. В этом случае фокус расположен в области  $D_1$ , в которой также размещен источник первичного поля  $\vec{E}_0, \vec{H}_0$ . В случае  $h > 2\Delta$  точка  $O_2$  является фиктивным фокусом, лежащим вне области  $D_1$ . Точка  $O_3$  во всех случаях является фиктивным фокусом. ■

### Заключение

Фокусирующие материальные параметры биизотропного экрана составляют два одномерных линейных многообразия, параметризуемых комплексной величиной  $\delta$ . Экран толщиной  $\Delta$ , выполненный из фокусирующего материала, осуществляет фокусировку поля магнитного диполя, расположенного перед экраном на расстоянии  $h$  от экрана. При условии  $0 < \alpha < 1$  ( $\alpha = \frac{h}{\Delta}$ ) образуются два фокуса – перед экраном и за экраном, т. е. имеем двухстороннюю фокусировку. Поле за экраном фокусируется в виде магнитного диполя, а поле перед экраном – в виде суммы магнитного и электрического диполей. При условии  $1 < \alpha < 2$  образуется один фокус перед экраном. При  $\alpha > 2$  фокусы отсутствуют.

### Список литературы

1. Виноградов, А.П. Электродинамика композитных материалов / А.П. Виноградов. – М. : Эдиториал УРСС, 2001. – 206 с.
2. Неганов, В.А. Современное состояние электродинамики искусственных киральных сред (обзор) / В.А. Неганов, О.В. Осипов // Физика волновых процессов и радиотехнические системы. – 2005. – Т. 8, № 1. – С. 7–33.
3. Гуляев, Ю.В. Метаматериалы: фундаментальные исследования и перспективы применения / Ю.В. Гуляев, А.Н. Лагарьков, С.А. Никитов // Вестник Российской академии наук. – 2008. – Т. 78, № 5. – С. 438–457.
4. Cui, T.J. Metamaterials. Theory. Design and Applications / T.J. Cui, D.R. Smith, R. Lui. – Springer, 2009. – 367 p.
5. Шевченко, В.В. Локализация стационарного электромагнитного поля при помощи плоской границы метаматериала / В.В. Шевченко // Успехи физических наук. – 2011. – Т. 181, № 11. – С. 1171–1182.
6. Проникновение электромагнитных волн через композитные экраны, содержащие идеально проводящие спирали / В.Т. Ерофеенко [и др.] // Инженерно-физический журнал. – 2011. – Т. 84, № 4. – С. 740–746.
7. Ерофеенко, В.Т. Численное исследование взаимодействия электромагнитных полей электрического и магнитного диполей с композитным экраном / В.Т. Ерофеенко, В.Ф. Бондаренко // Известия НАН Беларуси. Сер. физ.-техн. наук. – 2013. – № 4. – С. 113–120.
8. Ерофеенко, В.Т. Преобразование пучков электромагнитных волн при прохождении через экран из кирального метаматериала / В.Т. Ерофеенко, В.Ф. Бондаренко // Информатика. – 2013. – № 1. – С. 5–17.
9. Веселаго, В.Г. Электродинамика веществ с одновременно отрицательными значениями  $\epsilon$  и  $\mu$  // В.Г. Веселаго // Успехи физических наук. – 1967. – Т. 92, № 3. – С. 517–523.
10. Веселаго, В.Г. О формулировке принципа Ферма для света, распространяющегося в веществах с отрицательным преломлением // Успехи физических наук. – 2002. – Т. 172, № 10. – С. 1215–1218.
11. Ерофеенко, В.Т. Прохождение симметричных волн через бианизотропно-гиротропную перегородку в круглом волноводе с импедансными граничными условиями / В.Т. Ерофеенко, А.К. Сеницын // Информатика. – 2014. – № 4. – С. 5–18.
12. Шевченко, В.В. Геометрическая теория плоской линзы из кирального метаматериала / В.В. Шевченко // Радиотехника и электроника. – 2009. – Т. 54, № 6. – С. 696–700.
13. Ерофеенко, В.Т. Краевая задача проникновения электромагнитных полей дипольных источников через биизотропный экран / В.Т. Ерофеенко // Вестник БГУ. Сер. 1. – 2012. – № 2. – С. 71–76.

14. Ерофеенко, В.Т. Модели граничных условий на композиционных экранах для электромагнитных полей с осевой симметрией / В.Т. Ерофеенко // Известия НАН Беларуси. Сер. физ.-мат. наук. – 2010. – № 2. – С. 41–45.

15. Ерофеенко, В.Т. Аналитическое моделирование в электродинамике / В.Т. Ерофеенко, И.С. Козловская. – Минск : БГУ, 2010. – 304 с.

Поступила 28.11.2015

*Учреждение БГУ «НИИ прикладных проблем  
математики и информатики»,  
Минск, пр. Независимости, 4  
e-mail: bsu\_erofeenko@tut.by*

**V.T. Erofeenko**

### **TWO-SIDED FOCUSING OF ELECTROMAGNETIC FIELD OF POINT SOURCE BY MEANS OF PLANAR LENS FROM METAMATERIALS**

An analytical solution of the boundary-value problem of penetration of electromagnetic field of magnetic dipole through the planar lens from the focusing materials is suggested. It is demonstrated, that for the lenses from focusing materials two focuses are formed: first is over the screen and the second is in front of the screen in the source area. The field of focus in front of the lens is equal to the sum of the fields of magnetic and electric dipoles.

УДК 519.8

А.М. Бондоловский, М.Я. Ковалев

**ОПТИМИЗАЦИЯ ГИБКИХ ЦЕН ГОСТИНИЧНЫХ НОМЕРОВ**

*Предлагается подход к решению задачи отыскания гибких цен гостиницы, максимизирующих прибыль от сдачи номеров гостям. Подход включает определение входных параметров для последующего математического анализа, разделение спроса на несколько категорий, прогнозирование спроса, определение коэффициентов функции спроса и решение задачи математического программирования для оптимизации цен.*

**Введение**

Анализ литературы [1] показал, что большинство исследований по управлению доходами в гостиничном бизнесе посвящены либо прогнозированию спроса, либо оптимизации доходов при условии, что спрос (или его вероятностное распределение) изначально задан. Исследователи дорабатывают существующие методы прогноза и оптимизационные модели, разрабатывают новые, но изредка объединяют их в одну систему по управлению гибкими ценами гостиницы. В настоящей работе рассматриваются: решение задачи динамического ценообразования (Д-цена), включающей все указанные выше аспекты, и интеграция решения в реальные системы управления доходами гостиницы.

Задача Д-Цена – это динамическая и неопределенная задача нахождения цен на номера гостиницы при условии, что спрос чувствителен к цене, и использования критерия максимизации общей прибыли от сдачи номеров гостям. Практическая ситуация, описываемая как задача Д-Цена, возникает при бронировании номеров через интернет-сервис, который автоматически принимает заказы, если они могут быть выполнены.

**1. Задача Д-Цена**

Задача Д-Цена может быть сформулирована следующим образом. Существуют гостиничные номера нескольких типов и неопределенный спрос нескольких категорий. Категории характеризуются типом гостиничного номера, сезоном высокой или низкой загрузки, продолжительностью проживания гостей в номерах и прочими параметрами. Предполагается, что спрос зависит от цены так, что  $f_{\tau,c}(p_{\tau,c}) = a_{\tau,c} - b_c p_{\tau,c}$ , где категория  $c$  и день  $\tau$  заданы;  $f_{\tau,c}$  – функция спроса (число занятых номеров в категории  $c$  в день  $\tau$ );  $p_{\tau,c}$  – цена;  $b_c > 0$  – константа, называемая в экономической литературе (см., например, [2]) коэффициентом эластичности, которая отражает чувствительность спроса на номера к изменению цены, и  $a_{\tau,c}$  – константа. Предполагается, что известны прошлые значения спроса и цен, а также расходы на обслуживание каждого номера. Необходимо, основываясь на прошлых данных, определить коэффициенты  $b_c > 0$  и  $a_{\tau,c} > 0$  и найти такие цены, чтобы разница между общим доходом и общими расходами на обслуживание номеров в горизонте планирования была наибольшей, цены находились между установленными нижними и верхними границами и количество заказов не превышало емкости номерного фонда гостиницы.

Задача решается таким образом, что вначале посредством исторических методов прогноза определяются значения  $b_c$  и  $a_{\tau,c}$ . Далее полученные значения используются в задаче математического программирования с вогнутой квадратичной целевой функцией и линейными ограничениями, целью которой является максимизация общей прибыли гостиницы. Решением задачи будет нахождение оптимальных цен  $p_{\tau,c}^*$  для каждой категории  $c$  и дня  $\tau$  горизонта планирования. При известных оптимальных значениях цен можно подсчитать соответствующую

щий спрос  $a_{\tau,c} - b_{cp_{\tau,c}^*}$ . Эти значения спроса показывают загруженность гостиницы на каждый день горизонта планирования и могут быть использованы для планирования расписаний гостиничных служб.

Для решения задачи Д-Цена создано специальное программное обеспечение. Задача математического программирования решается стандартным оптимизационным решателем IBM ILOG CPLEX Optimization Studio Version 12.6 [3].

## 2. Общая схема решения задачи Д-Цена

Опишем подход к решению задачи Д-Цена на примере некоторой гостиницы Н. Подход включает в себя определение входных параметров для последующего математического анализа, разделение спроса на несколько категорий, прогнозирование спроса, определение коэффициентов функции спроса и решение задачи математического программирования для каждой категории и каждого дня горизонта планирования. Целевая функция задачи является квадратичной по причине перемножения цены и соответствующего значения спроса, линейно зависящего от цены. Максимизация прибыли для каждой категории и каждого дня ведет к максимизации прибыли во всей гостинице в горизонте планирования.

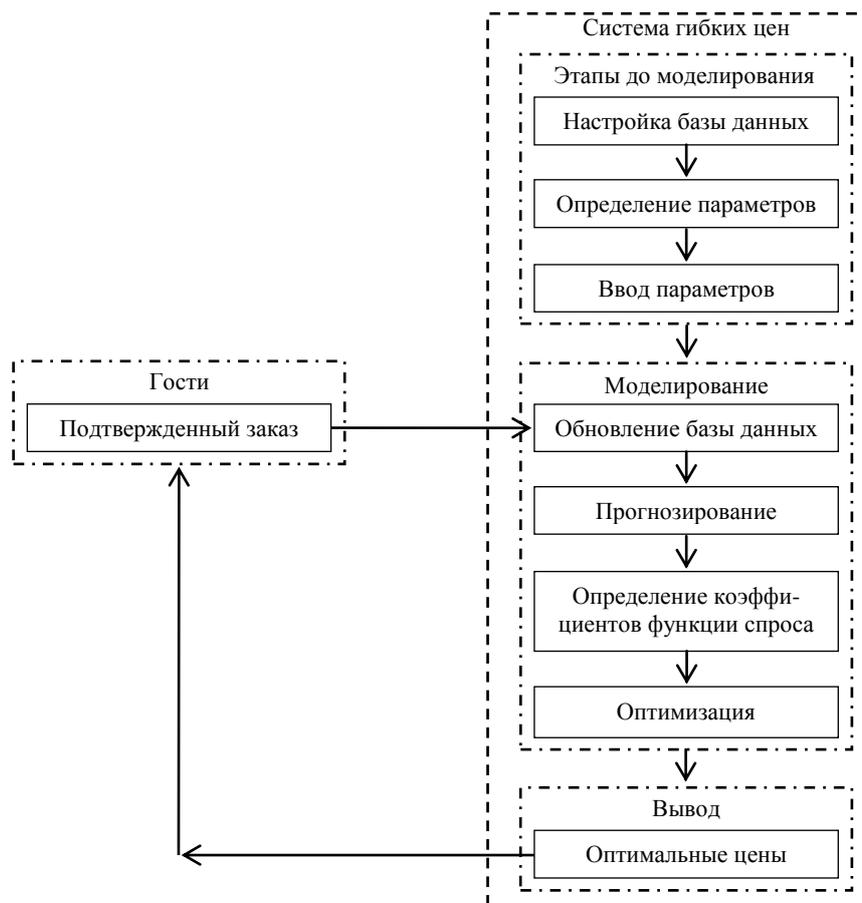
При решении задачи допустим, что разделение спроса на категории ведет к более точным результатам прогнозирования и оптимизации, спрос различных категорий не зависит друг от друга, а спрос эластичен и зависит только от цены.

Быстрое развитие информационных технологий и электронной коммерции поспособствовало росту внедрения систем гибкого ценообразования (СГЦ) в гостиницы и интереса ученых к этой области. Одно только условие, что большинство бронирований номеров поступает через Интернет, облегчает имитацию спроса, динамический пересчет, выставление подходящих цен и автоматическое распределение спроса по соответствующим категориям.

Предлагаемый подход к решению задачи Д-Цена также применяет преимущества современных информационных технологий. Приведем его основные этапы:

1. Настройка базы данных гостиницы для использования СГЦ.
2. Определение следующих входных параметров и переменных:
  - параметров, которые определяют категории спроса;
  - верхних и нижних границ на цены каждой категории и базовые цены;
  - операционных издержек обслуживания каждого типа номера;
  - горизонта планирования.
3. Ввод параметров в СГЦ.
4. Прогнозирование, которое включает:
  - формирование исторических периодов временных рядов количеств заселений в гостиницу для каждой категории;
  - формирование исторических периодов временных рядов продолжительности проживания, связанной с каждым заселением в каждой категории;
  - прогноз числа заселений для каждой категории спроса и дня горизонта планирования;
  - прогноз продолжительности проживания, связанной с каждым заселением, для каждой категории спроса и дня горизонта планирования;
  - арифметический подсчет прогнозируемого числа занятых номеров для каждой категории спроса и дня горизонта планирования.
5. Определение коэффициентов функций спроса путем вычисления:
  - коэффициента при переменной линейной функции спроса каждой категории;
  - константы функции спроса каждой категории.
6. Определение оптимальной цены для каждой категории и дня горизонта планирования.
7. Вывод следующих результатов:
  - оптимальной цены для каждой категории спроса и дня горизонта планирования;
  - ожидаемого количества занятых номеров в каждой категории спроса и дня горизонта планирования.

Этапы 4–7 повторяются после каждого подтвержденного заказа. На рисунке изображена схема взаимодействия процессов СГЦ.



Взаимодействие процессов СГЦ

Опишем подход к решению задачи Д-Цена и соответствующую СГЦ более подробно.

### 3. Разделение спроса, ввод параметров и переменных

Авторы предлагают разделять спрос на несколько категорий, которые характеризуются набором параметров. Запросы с одинаковым набором значений параметров относятся к одной категории. В исследовании Везефорда и др. [4] делается заключение, что разделение спроса на категории обеспечивает более высокую точность прогноза. Здесь следует отметить: хотя разделение спроса не создает трудностей в решении последующей оптимизационной задачи, оно может разрезать временные ряды спроса, тем самым уменьшая точность прогноза. Поэтому необходимо сохранять разумное соотношение между числом категорий и разреженностью данных в категориях.

Продemonстрируем подход к решению задачи Д-Цена на примере некоторой реальной гостиницы Н. Имеются данные о заказах и ценах этой гостиницы за 2009–2012 гг. Разделим спрос гостиницы Н на категории в соответствии со следующими параметрами: типом сезона (Сезон), днем недели (День), продолжительностью проживания (Длина), временем между датой совершения заказа и датой заселения (ДлинаДо) и классом номера (Класс).

Для гостиницы Н выделяются два сезона. Сезон низкой загрузки (Низкий) включает январь – март и ноябрь. Сезон высокой загрузки (Высокий) включает все месяцы года. Дни недели делятся на будни (Пн–Чт), которые состоят из всех дней с понедельника по четверг, и выходные (Пт–Вс), включающие дни с пятницы по воскресенье. Спрос внутри одного сезона и в один и тот же день недели не подвержен сезонным колебаниям, поэтому параметры Сезон и День позволяют избавиться от необходимости «очищать» спрос от сезонности.

Продолжительность проживания делится на два класса. В первом классе (7–) продолжительность меньше (или равна) семи дней, а во втором (8+) она больше семи дней. Время между датой совершения заказа и датой заселения может быть меньше (или равно) 7 (7–), 8–30 дней (8–30) и больше 30 дней (31+). Подход не рассматривает «заселение от стойки», т. е. без предварительного совершения заказа. В завершение выделяется пять классов номеров: эконом-класс (Э), эконом-плюс (Э+), бизнес-класс (Б), бизнес-плюс (Б+) и люкс (Л).

В результате вышеописанной классификации каждый заказ сопоставляется одной из 120 категорий спроса. Число 120 получено как произведение значений параметров категорий, т. е.  $120 = 2 \times 2 \times 2 \times 3 \times 5$ . Пусть  $c$  обозначает категорию спроса. Для категорий используются два типа обозначений:  $c = 1, \dots, 120$  и  $c = (\text{Сезон} = X, \text{День} = Y, \text{Длина} = Z, \text{ДлинаДо} = V, \text{Класс} = W)$ , где  $X \in \{\text{Низкий}, \text{Высокий}\}$ ,  $Y \in \{\text{Пн–Чт}, \text{Пт–Вс}\}$ ,  $Z \in \{7-, 8+\}$ ,  $V \in \{7-, 8-30, 31+\}$ ,  $W \in \{\text{Э}, \text{Э+}, \text{Б}, \text{Б+}, \text{Л}\}$ , например  $c = (\text{Сезон} = \text{Высокий}, \text{День} = \text{Пт–Вс}, \text{Длина} = 7-1, \text{ДлинаДо} = 7-, \text{Класс} = \text{Б})$ . Следует обратить внимание на то, что дата заселения в гостиницу однозначно определяет тип сезона и тип дня недели.

Другими входными параметрами являются горизонт планирования, издержки на обслуживание номера каждого типа, верхние и нижние границы значений цен в категориях спроса и базовая цена для каждой категории спроса. *Горизонт планирования* – это период времени в будущем, для каждого дня которого, начиная с текущего, необходимо определить оптимальные цены. Устанавливается условие, что цена не должна опускаться ниже издержек на обслуживание номера. Нижние и верхние границы на значения цен определяют диапазон, в котором может изменяться цена категории спроса. Базовая цена – эта актуальная конкурентная цена за номер в соответствующей категории спроса, на которую ориентируется менеджер гостиницы. Переменными в задаче являются цены на номера в каждой категории спроса и для каждого дня горизонта планирования. Отметим, что можно вводить либо верхние и нижние границы на значения цен в категориях спроса, либо базовую цену для каждой категории. Если менеджеру гостиницы привычнее работать с границами на значения цен в категориях спроса, то базовая цена рассчитывается как среднее значение этих границ. Если менеджер предпочитает базовую цену, которой может быть последняя реальная цена продажи в категории спроса, то нижние и верхние границы на значения цен могут отклоняться на заданное количество процентов от базовой цены.

Во избежание необходимости ручного ввода границ цен и базовой цены для каждой из 120 категорий менеджеру предлагается ввести границы цен или базовую цену только для одной категории каждого типа номера, а затем установить процентное отклонение от этих эталонных параметров для остальных категорий номера этого типа. Эталонная категория продается по самой высокой цене среди подмножества категорий одного типа номера. Для гостиницы  $N$  номера продаются по наибольшей цене в категориях (Сезон = Высокий, День = Пт–Вс, Длина = 7–, ДлинаДо = 7–, Класс =  $W$ ), где  $W \in \{\text{Э}, \text{Э+}, \text{Б}, \text{Б+}, \text{Л}\}$ . Границы цен или базовая цена для номеров категорий с параметрами Сезон = Низкий, День = Пн–Чт, Длина = 8+, ДлинаДо = 8–30 и ДлинаДо = 31+ будут автоматически уменьшаться на заданный процент, например на 15, 8, 8, 8 и 15 % соответственно. Менеджер гостиницы определяет и вводит значения этих процентных отклонений.

#### 4. Прогнозирование спроса

Для системы гибких цен выбор метода прогнозирования влияет на точность прогноза. Выбор конкретного метода прогнозирования может зависеть от исторических данных о заселениях и ценах и от специфики глобальной модели управления доходностью. Авторы предлагают использовать исторические методы прогноза, потому что разделение спроса на категории разрежает данные и усложняет отслеживание закономерностей, необходимых для более сложных методов прогноза.

Прогнозирование спроса начинается с формирования двух исторических временных рядов для каждой категории и дня горизонта планирования: количества совершенных заселений и продолжительности пребывания для каждого заселения. Далее спрогнозированное число бу-

дущих заселений и их продолжительности объединяются для получения будущего числа занятых номеров в каждой категории и каждом дне горизонта планирования. Эти числа определяются арифметическими подсчетами.

Менеджеры гостиниц планируют и устанавливают цены ежедневно, еженедельно, ежемесячно, раз в полгода и раз в год. Предлагаемая система способна находить оптимальные цены на периоды планирования до года. Один и тот же метод прогнозирования не может дать одинаковые результаты для краткосрочного, среднесрочного и долгосрочного горизонтов прогнозирования. То же самое верно и для разной степени плотности исторических данных категорий спроса. В результате для периодов планирования до трех месяцев система использует модифицированный метод двойного экспоненциального сглаживания Хольта либо модифицированный метод скользящего среднего, а от трех месяцев и более – модифицированный «метод такого же дня в прошлом году». Выбор метода прогнозирования в краткосрочном периоде до трех месяцев зависит от степени разреженности исторических данных категории спроса. Разреженные данные ведут к неточным оценкам сглаживающих коэффициентов уровня и тренда в модифицированном методе двойного экспоненциального сглаживания Хольта и, соответственно, к погрешности. Поэтому модифицированный метод скользящего среднего применяется для разреженных данных. Данные считаются разреженными, если в одной из категорий на период моделирования хотя бы в один из дней не произошло заселений. Точность экстраполяционных методов прогноза сильно уменьшается с увеличением горизонта планирования. Для увеличения точности прогноза в условиях разреженности спроса все три метода прогноза доработаны авторами.

Опишем каждый из применяемых методов прогноза.

Пусть  $d_i$  – наблюдаемое значение временного ряда в период времени  $i$ , т. е. число заселений в гостиницу в день исторического периода  $i$ ,  $i = 1, 2, \dots, t$ ;  $\hat{d}_i$  – спрогнозированное значение  $d_i$ , рассчитываемое для корректировки метода;  $e_i$  – ошибка прогноза для дня  $i$  и  $e_i = d_i - \hat{d}_i$ . Для оценки точности прогноза используется среднеквадратическая ошибка ( $MSE$ ):

$$MSE = (\sum e_i^2) / T,$$

где  $T$  – количество дней в прогнозе;  $t$  – последний день исторического периода, т. е. день осуществления прогноза;  $\hat{d}_{t'}$  – спрогнозированное число заселений на день  $t'$  горизонта планирования,  $t' = t + m$ ,  $m = 1, 2, \dots, T$ .

#### **4.1. Модифицированный метод двойного экспоненциального сглаживания Хольта (М-Хольт)**

Метод двойного экспоненциального сглаживания Хольта представляет собой расширенную версию простого экспоненциального сглаживания [5, 6]. Преимущество метода двойного экспоненциального сглаживания Хольта состоит в том, что он позволяет улавливать среднесрочный тренд временного ряда. Тренд учитывает как внутренние изменения в гостинице, так и внешние эффекты, влияющие на бизнес в регионе. Спрос, разбитый по категориям, имеет слабовыраженный тренд. Авторы пользуются методом Хольта, чтобы сохранить возможность влияния тренда. Метод Хольта является развитием метода экспоненциального сглаживания и дополняет сглаженный уровень временного ряда  $l_i$  трендом  $r_i$  [5]. Уровень и тренд рассчитываются по формулам

$$l_i = \alpha d_i + (1 - \alpha)(l_{i-1} + r_{i-1}), \quad r_i = \gamma (l_i - l_{i-1}) + (1 - \gamma)r_{i-1},$$

где  $\alpha$  – сглаживающий коэффициент ряда,  $0 \leq \alpha \leq 1$ ;  $\gamma$  – сглаживающий коэффициент тренда,

$$0 \leq \gamma \leq 1. \text{ Принято считать, что } l_1 = d_1, \text{ а } r_1 = \frac{(d_2 - d_1) + (d_3 - d_2) + (d_4 - d_3)}{3}.$$

Прогноз на день  $i + 1$  находится путем суммирования уровня и тренда:

$$\hat{d}_{i+1} = l_i + r_i.$$

При использовании метода коэффициентам  $\alpha$  и  $\gamma$  изначально присваиваем произвольные значения в рамках ограничений. Для дня  $i$  рассчитываем сглаженный уровень  $l_i$  и вычисляем ожидаемое значение тренда  $r_i$ . Рассчитав прогноз  $\hat{d}_{i+1}$ , находим ошибку прогноза  $e_i$ . После этого процесс повторяем для каждого из дней исторического периода до дня  $t$ . Минимизировав среднюю квадратичную ошибку, рассчитанную для всего исторического периода, подбираем оптимальные значения коэффициентов  $\alpha$  и  $\gamma$ . Определив оптимальные значения коэффициентов, делаем прогноз на дни горизонта планирования. Для прогноза на  $m$  дней вперед за пределы исторического периода применяем формулу

$$\hat{d}_{t+m} = l_t + mr_t,$$

где  $m$  – порядковый номер дня горизонта планирования,  $m = 1, 2, 3, \dots, T$ .

При использовании двойного экспоненциального сглаживания Хольта следует учитывать, что эффективная длина исторического периода зависит от конкретных данных и изменяется от одного месяца до одного года.

Для увеличения точности прогноза модифицируем метод двойного экспоненциального сглаживания Хольта. Оригинальному методу свойственно образовывать дробные значения заселений. Чтобы корректно учесть дробные части заселений и не исказить точность при округлениях, предлагается, начиная с первого дня горизонта планирования, суммировать дробные части числа заселений каждого дня. Как только сумма становится большей либо равной единице, появляется новое заселение. Случайным образом новое заселение добавляется к прогнозу одного из рассмотренных дней. Таким образом не теряется ни один заказ.

#### 4.2. Модифицированное скользящее среднее (М-Среднее)

Скользящее среднее – один из наиболее простых методов прогноза. Прогноз заселений на день  $t + 1$  определяется нахождением среднего значения числа заселений  $N$  предыдущих дней по формуле

$$\hat{d}_{t+1} = \frac{1}{N} \sum_{i=t-N+1}^t d_i,$$

где  $\hat{d}_{t+1}$  – прогноз количества заселений на день  $t+1$ ;  $d_i$  – количество заселений в день  $i$ ;  $N$  – количество предыдущих дней.

Критерием выбора числа  $N$  является наименьшая ошибка прогноза. В проведенных экспериментах она достигалась при  $N = 8$ . Прогноз числа заселений для каждого последующего дня, начиная с  $t + 2$ , равен значению в день  $t + 1$ .

Для скользящего среднего достаточно, чтобы исторический период каждой категории состоял из  $N$  дней. Метод скользящего среднего был доработан для учета дробных частей так же, как и предыдущий.

#### 4.3. Модифицированный метод «такого же дня в прошлом году» (М-Год)

Метод «такого же дня в прошлом году» для прогноза числа заселений на текущий год использует исторические данные прошлого года. Предполагается, что такой же день недели прошлого года способен правильно оценить порядок числа заселений в текущем году. Например, если в пятницу 48-й недели 2014 г. спрос составлял 23 номера, то в пятницу 48-й недели 2015 г. спрос, возможно, будет таким же.

Для увеличения точности и добавления динамики предлагается модифицировать метод следующим образом. Рассмотрим прогноз числа заселений на день  $t + 1$ . Допустим, что это пятница. Модифицированный метод «такого же дня в прошлом году» рассматривает число заселений этого же дня в прошлом году, но добавляет к нему среднее значение отклонений этого числа от количества заселений последних четырех пятниц. Если количество заселений в пятницу 48-й недели 2014 г. равно 23 и количества заселений в пятницы 47-й, 46-й, 45-й и 44-й не-

дель 2015 г. равны 25, 26, 23 и 24 соответственно, то спрогнозированное число заселений на пятницу 48-й недели 2015 г. будет равно  $24,5 = 23 + (2+3+0+1)/4$ . Учет дробных частей заселений происходит способом, аналогичным двум предыдущим методам прогноза.

Вышеперечисленные методы прогноза позволяют по каждой категории получать ежедневные количества заселений на определенный горизонт планирования. Однако специфика предлагаемой системы гибких цен подразумевает еще и необходимость прогноза продолжительности пребывания в гостинице для каждого заселения. Существующее разделение категорий по диапазонам продолжительностей установлено для возможности дифференциации цены и не детализирует продолжительность пребывания. Если расширить число значений параметра длительности пребывания до максимально возможного срока пребывания в гостинице, то необходимость в прогнозе продолжительности пребывания отпадает. Чтобы определить, на сколько ночей останутся гости по каждому заселению, используются приведенные выше методы прогнозирования. Кроме того, для каждой категории в дополнение к временному ряду с ежедневным количеством заселений формируется последовательность количеств проведенных ночей для каждого заселения. В табл. 1 показаны заселения в прошлом для некоторой категории, а в табл. 2 – продолжительность пребывания для каждого заселения. Согласно выбранному методу прогноза продолжительности заселения экстраполируются в будущее и последовательно присваиваются каждому спрогнозированному заселению.

Для получения будущих значений спроса исторические методы прогноза используют только свершившиеся заселения. Однако на момент осуществления прогноза могут существовать заказы на заселение в горизонте планирования, которые будут отменены до заселения. Чтобы спрогнозировать, какие заказы будут отменены, проводится эксперимент. Для каждого заказа случайным образом выбрасывается число от 0 до 1. Если случайное число меньше либо равно вероятности отмены заказа в рассматриваемой категории, то заказ отменяется. Вероятность отмены заказа в каждой категории рассчитывается из исторических данных и равна доли отмененных заказов от общего числа всех заказов. Рассматриваемая система гибких цен не требует проведения более точного эксперимента, показывающего, за сколько дней до заселения заказ будет отменен. Подтвержденные (без возможности отмены) заказы не добавляются к спрогнозированному числу заселений. Они уменьшают число свободных номеров в рассматриваемые дни заселения.

Таблица 1

Прогноз числа заселений

Дата	Кол-во заселений	Индексы заселений
26.09.2010	0	
01.10.2010	3	54, 55, 56
02.10.2010	0	
03.10.2010	0	
08.10.2010	0	
09.10.2010	1	57
10.10.2010	0	
15.10.2010	2	58, 59
16.10.2010	0	

Таблица 2

Прогноз продолжительностей заселений

Дата	Индексы заселений	Кол-во ночей
01.10.2010	54	2
01.10.2010	55	2
01.10.2010	56	2
09.10.2010	57	1
15.10.2010	58	2
15.10.2010	59	2

Результат прогнозирования является промежуточным. Для установления зависимости спроса от цены предполагается, что спрогнозированные заселения будут проданы по цене последней продажи на момент прогноза. Спрогнозированное число заселений и указанная цена соответствуют точке в прямоугольной системе координат «спрос – цена» для каждой категории и каждого дня горизонта планирования. Эта точка и рассчитанный по историческим данным коэффициент эластичности определяют конкретную линейно убывающую функцию зависимости спроса от цены. Далее эта функция используется в оптимизационной модели.

## 5. Оптимизация цен

Рассмотрим некоторую категорию спроса  $c$  и день горизонта планирования  $\tau$ . Предполагается, что спрос для этой категории и дня представляет собой линейную функцию  $f_{\tau,c}(p_{\tau,c}) = a_{\tau,c} - b_c p_{\tau,c}$ , где  $b_c > 0$  – коэффициент эластичности и  $a_{\tau,c} > 0$  – константа. Отметим, что в отличие от товаров первой необходимости и элитных товаров спрос на гостиничные номера чувствителен к цене и в большинстве случаев возрастает при ее снижении.

### 5.1. Коэффициент эластичности

Коэффициент эластичности  $b_c$  определяется методом простой линейной регрессии, который размещает прямую линию по отношению к заданному набору точек таким образом, чтобы сумма квадратов расстояний от точек до прямой была наименьшей. Для получения более точных расчетов коэффициента  $b_c$  важно использовать исторические данные о занятых номерах за весь имеющийся период. Независимой переменной функции регрессии является цена на номер, скорректированная с учетом инфляции, а зависимой – число номеров, проданных по этой цене. Построенная линия представляет собой историческую функцию спроса, и ее наклон определяет коэффициент эластичности  $b_c$ . Коэффициент эластичности подвержен изменению только в долгосрочном периоде под влиянием внешних экономических факторов. Следовательно, рассчитанный по историческим данным, он не зависит от  $\tau$  и одинаков для всех дней горизонта планирования одной категории. Напротив, константа  $a_{\tau,c}$  функции спроса  $f_{\tau,c}$  зависит и от категории спроса  $c$ , и от дня  $\tau$ .

При расчете коэффициента эластичности исторические цены подсчитываются с учетом инфляции следующим образом:

$$P_a = \frac{P_h}{1 + RI_x},$$

где  $P_a$  – цена, прошедшая инфляционную очистку;  $P_h$  – историческая цена без учета инфляции;  $RI_x$  – уровень инфляции в рассматриваемый месяц  $x$ . Уровень инфляции определяется по формуле  $RI_x = \frac{CPI_x - CPI_0}{CPI_0}$ , где  $CPI_x$  – индекс потребительских цен в рассматриваемый месяц, а  $CPI_0$  – индекс потребительских цен в базовый месяц, который является последним месяцем доступного исторического периода. В Республике Беларусь индексы потребительских цен публикуются на официальном сайте Национального статистического комитета [7].

Чрезмерная разреженность данных некоторых категорий и влияние непредсказуемых факторов могут вызывать образование отрицательных значений коэффициента эластичности  $b_c$ . В связи с предположением, что функция спроса не возрастает, отрицательные значения коэффициентов  $b_c$  считаются отклонением от закона. При возникновении подобных случаев  $b_c$  приравнивается к нулю.

Константа  $a_{\tau,c}$  рассчитывается по формуле  $a_{\tau,c} = \hat{z}_{\tau,c} + b_c p_c^0$ , где  $p_c^0$  – базовая цена категории  $c$  и  $\hat{z}_{\tau,c}$  – спрогнозированное число занятых номеров этой категории в день  $\tau$ .

### 5.2. Оптимизационная модель

Критерием оптимизационной модели служит максимизация общей прибыли гостиницы. Входными данными являются параметры  $a_{\tau,c}$  и  $b_c$  для каждой функции спроса  $f_{\tau,c}$ , нижние и верхние границы значений цен каждой категории  $c$ , издержки на обслуживание каждого типа номера и количество свободных номеров каждого типа в каждый день  $\tau$  горизонта планирования. Модель представляет собой задачу математического программирования

$$\max \sum_{c=1}^k \sum_{\tau=t+1}^{t+T} (a_{\tau,c} - b_c p_{\tau,c}) (p_{\tau,c} - h_c) - W \sum_{c=1}^k \sum_{\tau=t+1}^{t+T} y_{\tau,c} \quad (1)$$

при условии, что

$$L_{\tau,c} \leq p_{\tau,c}, \tau = t + 1, \dots, t + T, c = 1, \dots, k; \quad (2)$$

$$p_{\tau,c} \leq U_{\tau,c} + y_{\tau,c}, \tau = t + 1, \dots, t + T, c = 1, \dots, k; \quad (3)$$

$$a_{\tau,c} \geq b_c p_{\tau,c}, \tau = t + 1, \dots, t + T, c = 1, \dots, k; \quad (4)$$

$$p_{\tau,c} \geq h_c, \tau = t + 1, \dots, t + T, c = 1, \dots, k; \quad (5)$$

$$\sum_{c \in M_j} (a_{\tau,c} - b_c p_{\tau,c}) \leq R_{\tau,j}, \tau = t + 1, \dots, t + T, j = 1, \dots, 5; \quad (6)$$

$$p_{\tau,c_1} \leq p_{\tau,c_2}, c_1 \in M_1, c_2 \in M_2, \tau = t + 1, \dots, t + T; \quad (7)$$

$$p_{\tau,c_2} \leq p_{\tau,c_3}, c_2 \in M_2, c_3 \in M_3, \tau = t + 1, \dots, t + T; \quad (8)$$

$$p_{\tau,c_3} \leq p_{\tau,c_4}, c_3 \in M_3, c_4 \in M_4, \tau = t + 1, \dots, t + T; \quad (9)$$

$$p_{\tau,c_4} \leq p_{\tau,c_5}, c_4 \in M_4, c_5 \in M_5, \tau = t + 1, \dots, t + T; \quad (10)$$

$$p_{\tau,c} \geq 0, y_{\tau,c} \geq 0, \tau = t + 1, \dots, t + T, c = 1, \dots, k, \quad (11)$$

где  $p_{\tau,c}$  и  $y_{\tau,c}$  – переменные, а исходными данными являются:

$[t + 1, t + T]$  – горизонт планирования;

$L_{\tau,c}$  – нижняя граница на цену  $p_{\tau,c}$ ;

$U_{\tau,c}$  – верхняя граница на цену  $p_{\tau,c}$ ;

$h_c$  – издержки на обслуживание номера в категории  $c$ ;

$R_{\tau,j}$  – количество номеров типа  $j$ , свободных в день  $\tau$ ;

$y_{\tau,c}$  – вспомогательная переменная, которая позволяет нарушать верхние границы цен, когда эти границы делают область допустимых решений пустой;

$W$  – достаточно большое число, большее оптимального значения прибыли для задачи, в которой верхние границы цен отсутствуют. Например,  $W$  может быть равно сумме наибольших значений функций  $(a_{\tau,c} - b_c p_{\tau,c})(p_{\tau,c} - h_c)$  при условии, что существуют всего два ограничения  $p_{\tau,c} \geq h_c$  и  $a_{\tau,c} \geq b_c p_{\tau,c}$  по всем  $\tau$  и  $c$ . При таком  $W$ , если существует допустимое решение с  $y_{\tau,c} = 0$ , в любом оптимальном решении также будет выполняться равенство  $y_{\tau,c} = 0$  для всех  $\tau$  и  $c$ ;

$M_j$  – множество всех категорий, которые включают номера типа  $j$ . Предполагается, что множества  $M_j$  пронумерованы по неубыванию цен на номера.

Целевая функция (1) состоит из общей прибыли с положительным знаком и штрафов за расширение верхних границ цен с отрицательным знаком. Отметим, что штрафы за расширение границ являются искусственными и не влияют на реальную прибыль. Положительные значения переменных  $y$  указывают, насколько менеджеры гостиниц могут превысить верхние границы цен для уравнивания спроса и предложения в случаях чрезмерного спроса. Возможность превышать верхние границы цен позволяет менеджерам не отказывать гостям. Неравенства (2) и (3) определяют нижние и верхние границы цен соответственно. Ограничения (4) гарантируют, что спрос будет принимать только неотрицательные значения. Неравенства (5) требуют, чтобы цена не опускалась ниже издержек на обслуживание номеров. Ограничения (6) обеспечивают непревышение суммы запрашиваемых номеров каждого типа в разных категориях в день  $\tau$  над числом свободных номеров типа  $j$  в этот день. Ограничения (7)–(10) поддерживают иерархию цен между типами номеров.

Оптимизационная модель не допускает избыточного бронирования. Может случиться, что прогнозируемый при базовой цене спрос превысит доступное количество номеров. В этих случаях модель будет поднимать цены для сокращения спроса с тем, чтобы не нарушать ограничений на свободные ресурсы.

Выражения (1)–(11) – это задача математического программирования с вогнутой квадратичной целевой функцией и линейными ограничениями. Целевая функция вогнута, так как она

представляет собой сумму вогнутых квадратичных функций одной переменной. Задача может быть решена стандартным оптимизационным решателем, например IBM ILOG CPLEX Optimization Studio Version 12.6 [3].

Отметим, что задача (1)–(11) может быть разложена на  $T$  подзадач. Каждая подзадача рассматривает один день  $\tau$ ,  $\tau = t + 1, \dots, t + T$ . Оптимальное решение исходной задачи определяется оптимальными решениями подзадач.

Оптимизация – последний этап разрабатываемой системы динамического ценообразования. Решение задачи (1)–(11) определяет оптимальные цены  $p_{\tau,c}^*$  для каждой категории  $c$  и дня  $\tau$  горизонта планирования и оптимальные значения вспомогательных переменных  $y_{\tau,c}^*$ . Определив оптимальные цены  $p_{\tau,c}^*$ , можно рассчитать соответствующий спрос  $a_{\tau,c} - b_c p_{\tau,c}^*$ . Значения спроса могут быть использованы для оценки загрузки гостиницы и, соответственно, планирования деятельности обслуживающего персонала.

При возникновении ситуации, когда период пребывания заказа начинается в одной категории (сезон низкой загрузки или будни), а заканчивается в другой категории (сезон высокой загрузки или выходные), предлагается разделить период пребывания на несколько частей, каждая из которых соответствует одной категории, и рассчитывать цены для каждой части.

Решение задачи (1)–(11) может быть проанализировано и подтверждено или изменено лицами, принимающими решения. Подтвержденное решение передается на веб-сайт и становится доступным потенциальным гостям гостиницы.

Предложенное решение позволяет гостиницам реализовывать одну из двух стратегий принятий заказов. В соответствии с первой стратегией гостиница принимает любой заказ в случае наличия свободного номера. Вторая стратегия предписывает принимать заказы в каждой категории до оптимального прогнозного количества  $a_{\tau,c} - b_c p_{\tau,c}^*$ . Заказы, превышающие оптимальное количество, должны отклоняться. Эффективность второй стратегии сильно зависит от точности прогноза. Независимо от выбора стратегии предлагается обновлять цены после каждого совершенного (оплаченного) заказа. Также предлагается установить несколько горизонтов планирования разной продолжительности, например 1, 7, 31, 90, 180 и 360 дней, и обновлять цены для всех них одновременно. Увеличение продолжительности горизонта планирования сокращает точность прогноза и соответственно качество решений. В связи с этим решения для коротких горизонтов планирования более приоритетны по сравнению с решениями для долгосрочных периодов.

## 6. Компьютерный эксперимент

Для оценки эффективности предлагаемой системы гибких цен проведен компьютерный эксперимент, целью которого является сравнение реального дохода гостиницы  $N$  в прошлом периоде со смоделированным потенциальным доходом, сгенерированным системой за тот же прошлый период, который называется *сравнительным периодом*, при условии, что зависимость спроса от цены описывается функциями  $f_{\tau,c}$ .

При проведении эксперимента использовалась конфиденциальная финансовая информация. Поэтому все входные числовые значения были умножены на один и тот же коэффициент и округлены. Издержки на обслуживание номеров  $h_c$  равны 50 для всех типов номеров. Базовые цены наиболее дорогих категорий номеров классов Э, Э+, Б, Б+ и Л равны 139, 149, 159, 175 и 189 соответственно. Процентные отклонения для категорий  $c$  параметрами Сезон = Низкий, День = Пн–Чт, Длина = 8+, ДлинаДо = 8–30 и ДлинаДо = 31 + равны 20, 10, 10, 10 и 15 % соответственно. В эксперименте использован метод прогноза М-Среднее.

Период в прошлом равен 90 дням. Закрепленный день  $t$  разделит его на два периода: первоначальный исторический период, равный 30 дням и включающий дни  $t-29, t-28, \dots, t$ , и первоначальный горизонт планирования, равный 60 дням и включающий дни  $t+1, t+2, \dots, t+60$ . Для сравнительного периода были выбраны 14 дней  $t+31, t+32, \dots, t+44$  по той причине, что ни один заказ, осуществленный в день  $t$  и обладающий параметром ДлинаДо = 31+, не может быть реализован в день  $t+30$  или раньше.

...Таблица 3

Сравнение доходов

Сценарий «низкий рост»			Сценарий «высокий рост»			Сценарий «устойчивый»					
Но- мер	Дни сравни- тельного периода	Реальный доход	Смоделиро- ванный доход	Но- мер	Дни сравни- тельного периода	Реальный доход	Смоделиро- ванный доход	Но- мер	Дни сравни- тельного периода	Реальный доход	Смоделиро- ванный доход
1	11.09.11	142,5	1441,6	1	05.03.12	0	162,9	1	15.07.11	0	1561,2
2	12.09.11	0	624,7	2	06.03.12	285	162,2	2	16.07.11	527,5	1472,0
3	13.09.11	0	261,0	3	07.03.12	0	149,0	3	17.07.11	325	979,2
4	14.09.11	487,5	606,2	4	08.03.12	142,5	190,9	4	18.07.11	152,5	829,6
5	15.09.11	782,5	471,7	5	09.03.12	557,5	1404,1	5	19.07.11	822,5	108,0
6	16.09.11	1907,5	382,2	6	10.03.12	1167,5	474,3	6	20.07.11	162,5	447,6
7	17.09.11	507,5	1774,4	7	11.03.12	0	172,8	7	21.07.11	842,5	100,1
8	18.09.11	477,5	1701,8	8	12.03.12	325	135,2	8	22.07.11	822,5	411,5
9	19.09.11	0	745,2	9	13.03.12	152,5	449,2	9	23.07.11	822,5	468,0
10	20.09.11	0	556,2	10	14.03.12	152,5	337,3	10	24.07.11	182,5	1146,0
11	21.09.11	0	878,4	11	15.03.12	162,5	212,2	11	25.07.11	0	591,4
12	22.09.11	0	938,6	12	16.03.12	945	710,3	12	26.07.11	0	463,5
13	23.09.11	3025	1034,6	13	17.03.12	872,5	731,8	13	27.07.11	0	423,2
14	24.09.11	1270	608,8	14	18.03.12	0	663,9	14	28.07.11	152,5	663,3
Общий доход за период		8600,0	12 025,4	Общий доход за период		4762,5	5956,1	Общий доход за период		4812,5	9664,7
<b>Рост</b>			<b>39,8%</b>	<b>Рост</b>			<b>25,0%</b>	<b>Рост</b>			<b>100,8%</b>
				<b>Средний рост</b>							<b>55,2%</b>

Эксперимент проводился на ПК с процессором Intel Core i5×2 ГГц и 4 ГБ оперативной памяти в операционной системе MS Windows 8.1 Pro (64 bit). Для решения задачи математического программирования (1)–(11) использовался решатель IBM ILOG CPLEX Optimization Studio Version 12.6 [3]. Были рассмотрены три сценария. Каждый сценарий характеризовался разной скоростью и объемом изменения спроса в горизонте планирования по сравнению с историческим периодом и, соответственно, разными прошлыми периодами по 90 дней. В сценарии «устойчивый» ежедневное количество заселений в историческом периоде мало отличается от ежедневного количества заселений в горизонте планирования. Сценарии «медленный рост» и «быстрый рост» характеризуются медленным и быстрым ростом ежедневного количества заселений в горизонте планирования по сравнению с ежедневными количествами заселений в историческом периоде.

В каждом сценарии алгоритм запускался для  $\tau = t, t + 1, t + 13$ . При каждом запуске система пользовалась историческим периодом  $\tau - 29, \tau - 28, \dots, \tau$  и определяла цены на горизонт планирования  $\tau + 1, \tau + 2, \dots, \tau + 60$ . Смоделированный доход рассчитывался для дня сравнительного периода  $\tau + 31$ . Результаты эксперимента приведены в табл. 3.

Эксперимент продемонстрировал, что предлагаемая система гибких цен увеличивает средний доход гостиницы по сравнению с доходами от продаж при статичной политике ценообразования. Кроме того, благодаря тому что структура издержек гостиницы характеризуется высокими постоянными и низкими переменными издержками, увеличение доходов вносит относительно больший вклад в прибыль, чем в доход. При более низких переменных издержках вклад в общую прибыль будет выше.

Нулевые значения реального дохода в дни сравнительных периодов объясняются отсутствием продаж номеров в эти дни. Система гибких цен показала наибольший рост дохода в сценарии «устойчивый», при котором значения спроса в горизонте планирования слабо отличались от значений в историческом периоде. Высокая точность прогноза внесла значительный вклад в общий рост дохода. Скорость изменения будущего спроса уменьшила точность прогноза в сценариях «низкий рост» и «высокий рост», однако не повлияла на положительное увеличение дохода.

### Заключение

В статье описан подход к решению задачи Д-Цена и соответствующая система определения гибких цен гостиницы. Подход предполагает разделение спроса гостиницы на несколько категорий, прогнозирование значений коэффициентов функции спроса и решение задачи математического программирования с вогнутой квадратичной целевой функцией и линейными ограничениями для отыскания оптимальных цен для каждой категории и дня горизонта планирования.

С точки зрения объема требуемых вычислительных ресурсов (времени работы процессора и памяти) предлагаемый подход является более предпочтительным, чем существующие методы стохастического программирования, динамического программирования и нечеткого математического программирования решения задач определения гибких цен.

### Список литературы

1. Бондоловский, А.М. Обзор моделей управления доходностью в гостиничном бизнесе / А.М. Бондоловский // Информатика. – 2014. – № 10. – С. 66–83.
2. Houthakker, H. Harvard economic studies / H. Houthakker, L. Taylor. – Harvard : Harvard University Press, 1970. – Vol. 126. – 321 p.
3. IBM. ILOG CPLEX Optimization Studio. IBM [Electronic resource]. – 2015. – Mode of access : <http://goo.gl/9Uy8PA>. – Date of access : 10.12.2015.
4. Weatherford, L.R. Forecasting for hotel revenue management: testing aggregation against disaggregation / L.R. Weatherford, S.E. Kimes, D.A. Scott // Cornell Hotel and Restaurant Administration Quarterly. – 2001. – Vol. 42. – P. 53–64.

5. Gardner, E. Exponential smoothing: The state of the art – Part ii / E. Gardner // Intern. J. of Forecasting. – 2006. – Vol. 22, no. 4. – P. 637–666.
6. Forecasting uncertain hotel room demand / M. Rajopadhye [et al.] // Information Sciences. – 2001. – Vol. 132, no. 1–4. – P. 1–11.
7. Национальный статистический комитет Республики Беларусь [Электронный ресурс]. – 2015. – Режим доступа : [http://belstat.gov.by/ofitsialnaya-statistika/otrasli-statistiki/tseny/operativnaya-informatsiya\\_4/](http://belstat.gov.by/ofitsialnaya-statistika/otrasli-statistiki/tseny/operativnaya-informatsiya_4/). – Дата доступа : 11.09.2015.

Поступила 02.02.2016

*Объединенный институт проблем  
информатики НАН Беларуси,  
Минск, Сурганова, 6  
e-mail: andrei.bandalouski@gmail.com*

**A.M. Bandalouski, M.Y. Kovalyov**

### **OPTIMIZING HOTEL DYNAMIC PRICES**

An approach to solve a problem of determining optimal dynamic prices for hotel rooms is suggested. It includes selection of input parameters for the succeeding mathematical analysis, disaggregation of the demand into several categories, demand forecasting, simulation of demand-price relations, and a mathematical programming model for price optimization.

УДК 65.011.56

А.В. Пузанов

## ИНФОРМАЦИОННАЯ ПОДДЕРЖКА ЖИЗНЕННОГО ЦИКЛА ГИДРОПРИВОДОВ

*Рассматриваются направления повышения эффективности производства гидроприводов посредством применения информационной поддержки жизненного цикла изделия. Приводятся примеры использования программных комплексов для аналитических исследований на различных этапах жизненного цикла гидроприводов.*

### Введение

Потребители продукции машиностроения требуют от гидроприводостроения повышения технических, эксплуатационных и мощностных характеристик. Предприятиям машиностроения (в том числе гидроприводостроения) принято ориентироваться на тенденции развития ВПК, поскольку он был и остается локомотивом машиностроительной отрасли. От предприятий ВПК требуется повышение объемов производства на 15 % при сохранении численности работников, фактически на 15 % обязана повыситься и производительность труда. Вариантов решения задачи в данной постановке не столь много, поскольку повысить объемы производства на оборудовании без нарушения техпроцесса и перегрузки оборудования проблематично. Именно поэтому решение этой задачи видится в комплексном повышении эффективности бизнес-процессов на уровне организационных мероприятий, приоритета использования информационных технологий и высокопроизводительной вычислительной техники, поскольку прогресс и темпы роста данных направлений очевидны и намного превышают требуемые целевые показатели.

В условиях сокращения бюджета от организаций требуются дополнительные усилия по управлению себестоимостью, сроками изготовления и соблюдению все более жестких требований к изделиям. Решая данные задачи, предприятия формируют бизнес-процессы в едином информационном пространстве, реализуя тем самым информационную поддержку жизненного цикла изделий (ЖЦИ).

Задачи структурного исследования информационной системы на предмет ее эффективности или для изучения вариантов ее трансформации связаны с изучением разнородных бизнес-процессов, описывающих ЖЦИ. Для этих задач применяется методология и инструменты функционального моделирования (по стандарту IDEF), позволяющие исследовать структуру, параметры и характеристики процессов в производственно-технических и организационно-экономических системах. В российском правовом поле методология функционального моделирования регламентирована ГОСТ Р 50.1.028–2001.

Моделирование процессов в идеологии IDEF0 начинается с построения основной диаграммы А0, которая описывает функционирование конкретного (в нашем случае машиностроительного) предприятия в целом. Далее основная диаграмма разворачивается на составляющие диаграммы А1–А9 (рис. 1), которые, в свою очередь, декомпозируются до элементарного функционала рабочего места или функции работника.

Исследование функциональной модели позволяет проводить анализ изменений в бизнес-процессах и обосновывать применение средств и инструментов информационной поддержки ЖЦИ. В частности, на данной модели было проведено изучение информационных потоков для корректного использования аналитических инструментов поддержки конструкторско-технологических, экономических и тому подобных решений.

Поскольку концепцией информационной поддержки ЖЦИ подразумевается виртуализация реальных процессов (организационного, технического, информационного и др.), их анализ и моделирование образуют отдельный пласт задач и собственное информационное пространство со своими системами поддержки жизненного цикла виртуальных моделей изделий.

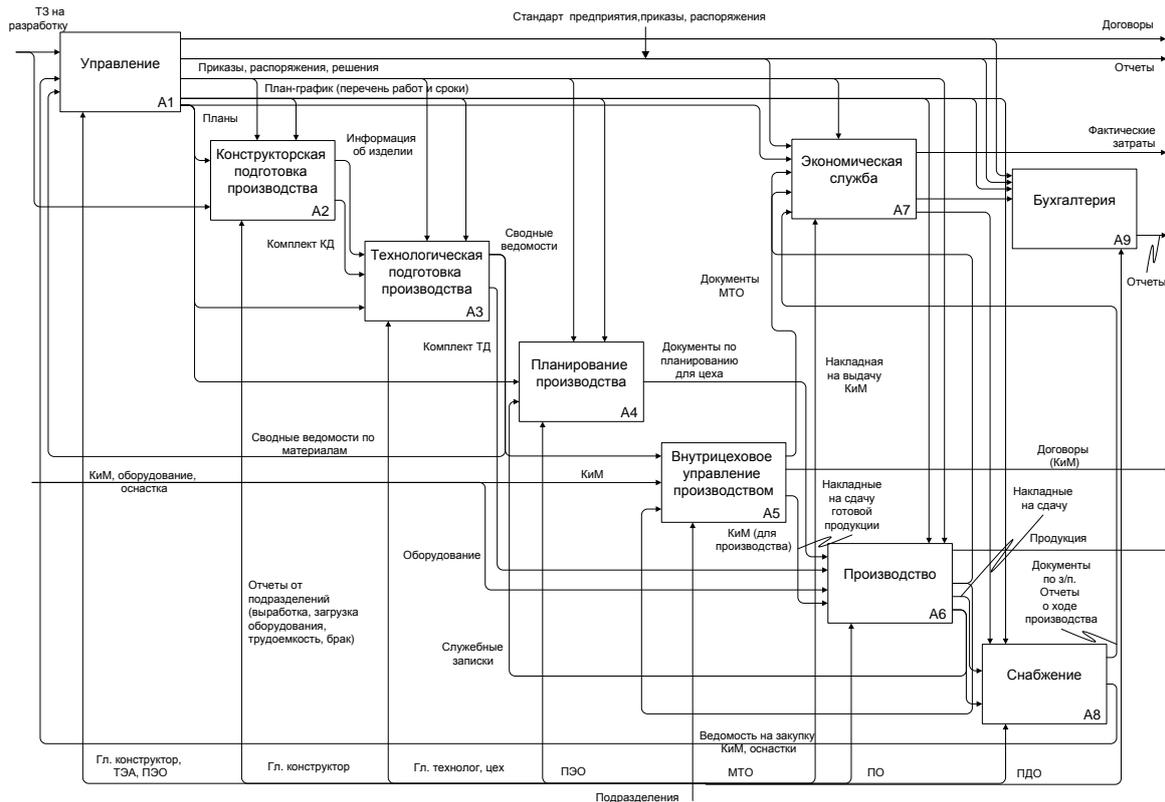


Рис. 1. Функциональная схема машиностроительного предприятия

Обычно концепция ЖЦИ на предприятиях реализуется следующим образом: на этапе проектирования создается комплект конструкторской документации, к которой на каждом последующем этапе ЖЦИ добавляется дополнительная информация, в лучшем случае – как подтверждение конструктивных, технологических и прочих решений в виде документов в архиве (в электронном виде, в общедоступном формате).

В настоящей работе рассматривается использование аналитических средств в качестве основного инструмента сопровождения ЖЦИ. Инструменты аналитических исследований являются частью ЖЦИ (CAE-анализ), но в контексте данной работы рассматриваются как средства системы поддержки принятия решений (оценки адекватности решений при формировании базы знаний) для других этапов ЖЦИ.

## 1. Использование средств инженерного анализа для поддержки жизненного цикла гидроприводов

Преимущества использования аналитических средств очевидны: если локальные компьютерные модели позволяют уменьшить время работы над изделием на каждом отдельном этапе его жизненного цикла, то их комплексный анализ сокращает количество фальш-итерационных процессов при проверке изделия на соответствие требованиям ТЗ (технологичности, функциональности и эргономичности).

Приведенные ниже иллюстрации отражают работу, выполняемую специалистами ОАО «Специальное конструкторское бюро приборостроения и автоматики» (г. Ковров, Россия) в рамках конструкторско-технологической подготовки производства гидроприводов, а также результаты совместных работ со смежными предприятиями и перспективные направления реализации концепции ЖЦИ.

В качестве инструментов аналитического сопровождения ЖЦИ использовались программные комплексы мультифизического анализа Autodesk Simulation Multiphysics, MSC.AFEA, Flow-3D, а также программный комплекс автоматизации интегрированных этапов подготовки производства «Интермех».

Рассмотрим возможности аналитических средств для решения задач на всех этапах ЖЦИ:

1. *На этапе маркетинговых исследований.* Для изделий гидроприводостроения характерна непрезентабельность внешнего вида и внутреннего устройства. С целью усиления привлекательности и наглядности принципов работы устройств используются материалы моделирования внутренних течений рабочей жидкости или тепловых потоков (рис. 2).

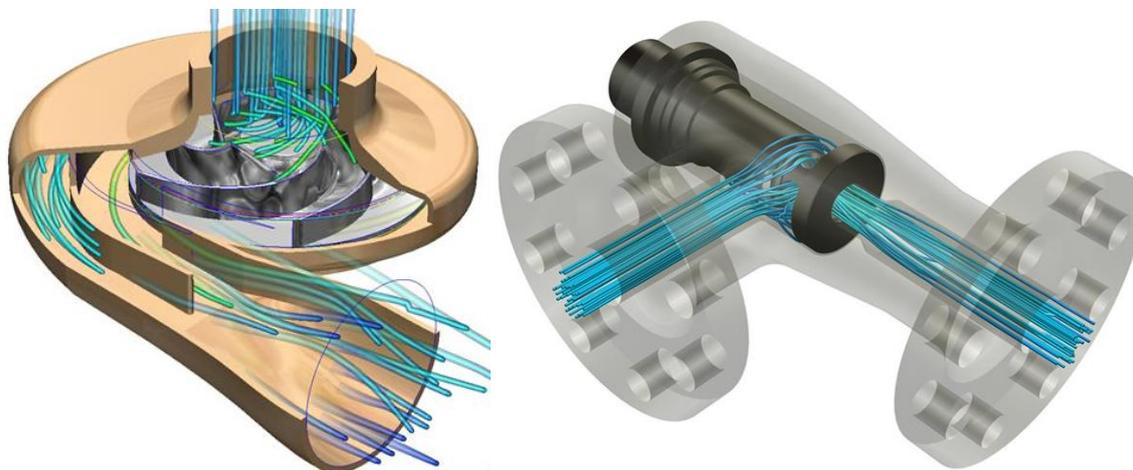


Рис. 2. Демонстрация принципа работы изделия

Отдельным направлением использования средств САЕ является производство рекламной продукции (анимации) с красочным детальным представлением рабочих процессов изделия (обычно внутренних, скрытых от внешнего взгляда).

Результаты моделирования на данном этапе ЖЦИ используются также в качестве наглядных инструментов, поясняющих отличительные преимущества конкретного продукта или отдельных конструкторско-технологических решений.

2. *На этапе проектирования.* Продукция гидроприводостроения отличается высокой вариативностью конструкций, жесткими допусками и разнофизичным характером внутренних рабочих процессов и внешних факторов. Чтобы минимизировать число дорогостоящих изменений, вносимых на поздних этапах разработки, конструкторам необходимо учитывать соображения технологичности деталей изделий уже на этапе проектирования.

Интегрированные средства проектирования позволяют предприятиям сокращать себестоимость и сроки выпуска продукции. Это повышает степень повторного использования и адаптации проектных решений, что, в свою очередь, также снижает себестоимость. Поддержка инноваций обеспечивается на основе постоянной совместной работы сотрудников в общем информационном пространстве.

Для этапа конструкторской проработки существует наибольшее количество методических работ и опытной реализации использования данных инструментов.

Методики применения аналитических средств и информационных технологий формализованы и методологизированы до мелочей, включая многоуровневые модели, схемы, процедуры оптимизации и мультидисциплинарность.

Применительно к гидроприводостроению методы автоматизации моделирования приводов и их элементов нашли отражение в работах И.П. Норенкова, Я.А. Даршта, А.В. Пузанова, Д.В. Багаева, С.А. Воронова и др. [4, 6, 10–12] с различной детализацией области и объектов исследований. Суть работ заключается в переносе исследовательской составляющей процесса конструирования в область математического моделирования.

В работах Я.А. Даршта рассматривается методология математического моделирования рабочих процессов в элементах гидроприводов (в гидромашинах и гидроаппаратуре), доведенная до инженерных методик, позволяющих в полуавтоматизированном режиме получать конструктивные элементы, отвечающие требованиям ТЗ [10].

В работах С.А. Воронова дана методология процессов проектирования гидромашин [11].

В предыдущих работах автора предлагается комплексный подход к моделированию рабочих процессов в элементах гидроприводов для гидромашин и гидроаппаратов, учитывающий мультифизичный характер этих процессов и их взаимовлияние [6]. Так, на рис. 3 представлена совмещенная картина гидродинамики в зоне всасывания в камеру блока цилиндров рабочей жидкости и напряженно-деформированного состояния поршня при контакте с гильзой блока цилиндров. Данный подход позволяет прогнозировать (планировать) ресурс изделия до начала его производства и эксплуатации.

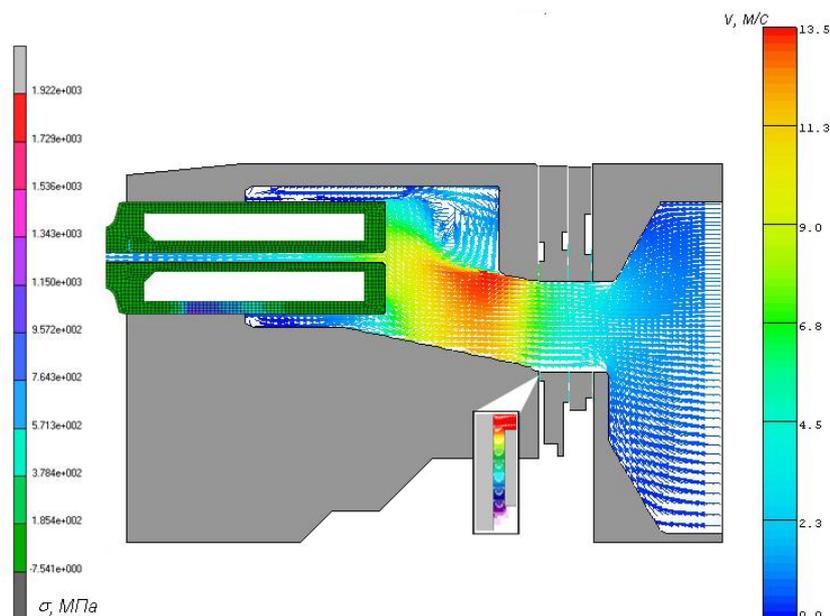


Рис. 3. Результат мультидисциплинарного анализа ходовой части гидромашин

В указанных работах рассматривается детализация процесса конструкторской проработки изделия на микро-, макро- и метаяуровнях. Они ориентированы на различных потребителей моделей: от инженера со средней квалификацией до аналитика и инженера-исследователя, владеющих должными знаниями для проведения процедур многокритериальной оптимизации конструкции гидропривода.

В процессе работ по моделированию и аналитическим исследованиям проводится либо наработка базы знания (для экспертной системы или системы принятия решений), либо построение инженерных методик (на основе полученных коэффициентов и зависимостей), либо углубленная детализация конструкторских решений.

3. *На этапе технологической подготовки производства.* Подход, основанный на использовании 3D-моделей деталей и сборок, позволяет устранить необходимость выпуска дополнительной документации с простановкой технологических параметров путем встраивания технических требований непосредственно в 3D-модель. Product Manufacturing Information (PMI) обеспечивает запись технических требований в 3D-модель (ГОСТ 2.052–2013), что в полной мере помогает раскрыть замысел конструктора, устранить необходимость в использовании 2D-чертежей и гарантирует, что готовое изделие будет соответствовать своему функциональному назначению. Данный подход относится к организационным мероприятиям, призванным сократить сроки конструкторско-технологической подготовки производства гидроприводов и их элементов.

Встроенные в 3D-модель данные PMI оптимизируют целый ряд производственных процессов. Такие данные применяются при автоматическом анализе размерных цепей и оценке наихудшего варианта сочетания допусков при сборке. Кроме того, данные PMI используются при разработке документации на импортированные модели деталей и узлов. К PMI можно применить логические ограничения, что позволяет интеллектуально реагировать на изменения в представлении основной модели изделия и появление его новых исполнений или новых ва-

риантов конструкции отдельных элементов. Кроме того, эта информация обеспечивает автоматическую генерацию управляющих программ для станков с ЧПУ и координатно-измерительных машин.

Большинство работ по использованию аналитических инструментов для технологической подготовки производства изделий посвящено отработке на технологичность и моделированию технологических процессов для обеспечения заданного качества продукции. В частности, средства инженерного анализа применяются для моделирования процессов обработки материалов резанием в качестве экспресс-оценки эффективности используемых режущих инструментов [7]. На рис. 4, *а* изображена форма режущей пластины по ГОСТ 17163-90, на рис. 4, *в* – форма режущей пластины фирмы Sandvick, соответствующие формы стружки при обработке этим инструментом – на рис. 4, *б* и *г*.

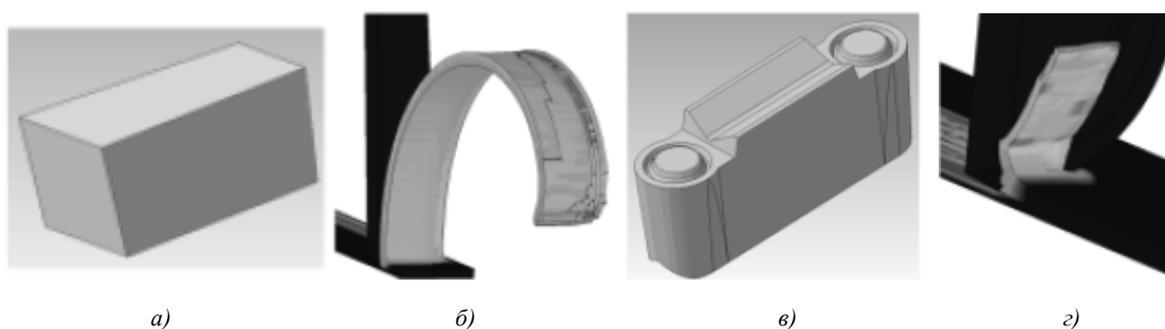


Рис. 4. Форма режущего инструмента и форма стружки при обработке

Приведенные выше результаты позволяют построить график зависимости формы и величины стружки от параметров металлообработки, что дает возможность исключить ситуации наматывания стружки на деталь и инструмент во время обработки.

Кроме того, проводится анализ жесткости всей системы СПИД (станок – приспособление – инструмент – деталь) [8] и деформации инструмента в частности (рис. 5). Эти результаты позволяют обоснованно подбирать инструмент и проектировать оснастку, обеспечивающие заданные конструктором допуски на размеры и форму деталей.

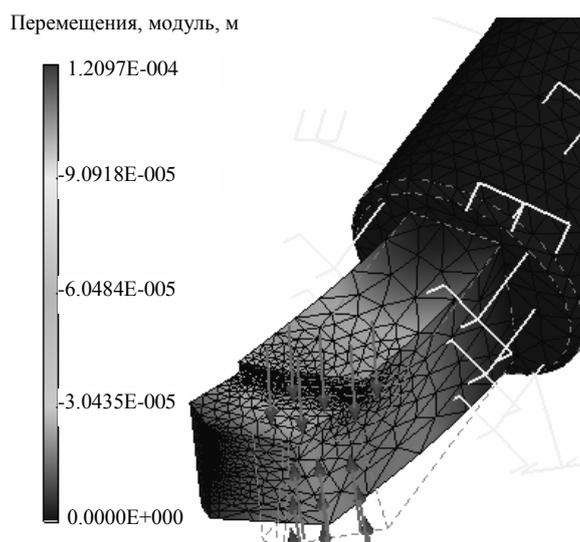


Рис. 5. Результаты анализа деформации инструмента

Для технологических процессов повышенной сложности (например, растачивания глубоких отверстий на токарных станках) выполняется анализ теплообмена, протекающего между рабочим инструментом, смазочно-охлаждающей жидкостью (СОЖ) и обрабатываемой деталью

с целью определения гидродинамических параметров, оптимальных для штатного функционирования инструмента и качественной обработки деталей [9]. На рис. 6 показано распределение температуры по детали, инструменту и СОЖ.

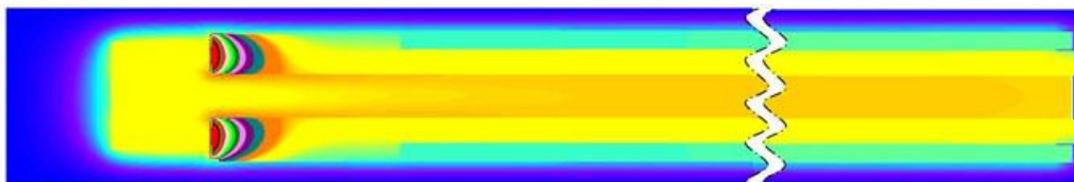


Рис. 6. Поля температуры в плоскости симметрии модели при температуре резания 600 °С и температуре подаваемой СОЖ 35° С

Последней тенденцией моделирования процесса обработки детали стало определение износа инструмента в зависимости от режимов резания (рис. 7).

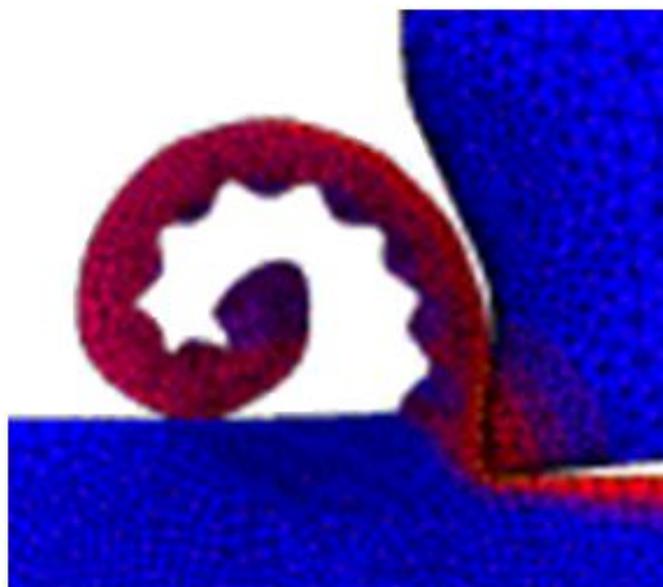


Рис. 7. Моделирование процесса резания

Аналитические средства анализа технологических процессов позволяют исследовать широкий спектр стратегий многокоординатной обработки сложных поверхностей (включая параллельную обработку), полностью устранить риск столкновения инструмента, обосновать изменение размеров при обработке, обеспечить параметры контроля и управления оборудованием.

На этапе технологической подготовки осуществляются также работы конструкторского характера: проектирование оснастки, приспособлений, производственных линий и участков. В решении данных задач используются инструменты аналитических исследований аналогично описанному выше этапу проектирования.

Применительно к производственному комплексу гидроприводов реализация технологии адаптивной обработки изделия требует наличия зависимостей усилий и напряженно-деформированного состояния инструмента и детали от параметров и режимов резания.

4. *На этапе производства.* Модельные исследования проводятся на макро- и метаяровнях для планирования производственной загрузки, логистики, на микроуровне – для моделирования электропитания оборудования, систем охлаждения и вентиляции производственных помещений или для моделирования загрязнений выбросами (рис. 8), а также иных катастроф техногенного и природного характера, вероятных для данного типа или места производства. Отсутствие виртуальных (математических) моделей затрудняет поиск проблемных мест, причин отказов и т. п.

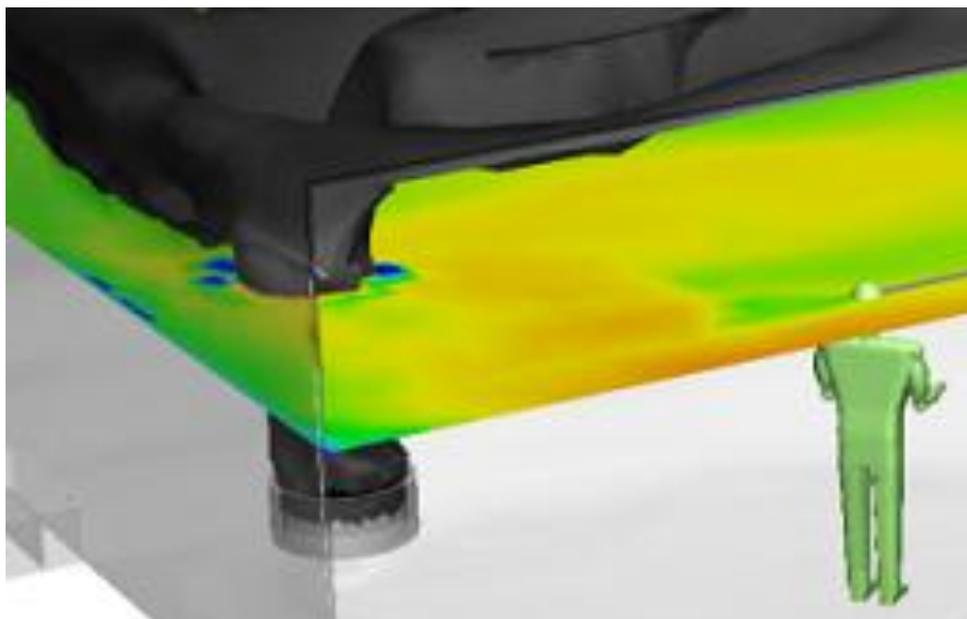


Рис. 8. Моделирование загрязнений от продуктов производственного процесса

Использование моделирования на этапе контроля обеспечивает анализ изменения допусков при изменении температуры.

5. *На этапе испытаний.* Модельные исследования проводятся для процессов, трудно-реализуемых на практике либо аварийно-экстремального характера (взрыв, краш-тест и т. п.), в основном для сертификации продукции на соответствие отраслевым нормам и правилам.

На рис. 9 показаны распределительный диск гидромашины, разрушенный в результате испытаний, и результат моделирования, призванного определить значения внешних факторов воздействия, при которых произошло данное событие [6].

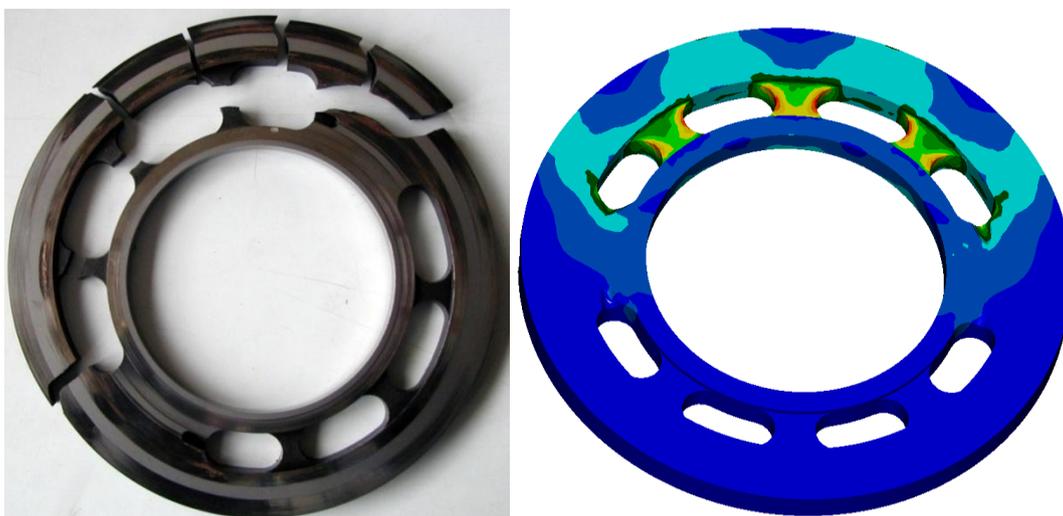


Рис. 9. Анализ причин разрушения детали гидромашины – распределителя стального

Другим направлением в использовании аналитических моделей на этапе испытаний является моделирование ускоренных ресурсных испытаний. Самым важным моментом на данном этапе ЖЦИ является формирование обратных связей с моделями этапов согласования технического задания, конструкторской проработки вариантов изделия и отработки технологичности его производства для верификации принятых допущений и оценки точности и адекватности моделей и конструктивных решений.

6. *На этапе эксплуатации.* Исследования поведения изделия на этапе эксплуатации проводятся, как правило, для анализа причин отказов (штатная нагрузка, перегрузка, недокументированное использование и т. п.) с целью определения ответственного или виновного в аварии, а также для анализа долговечности (рис. 10) и определения остаточного ресурса при повторном использовании изделия, бывшего в эксплуатации [13].

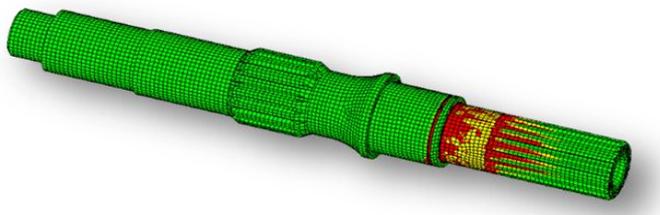


Рис. 10. Результат моделирования запаса надежности (по критерию усталости) вала гидромашин при количестве циклов, соответствующем 3000 ч работы

Применение средств моделирования для данного этапа жизненного цикла (обслуживание, текущий и капитальный ремонт) обосновывает сроки межремонтного обслуживания, обеспечивая высокую эффективность и бережливость данного этапа, а также сокращение расходов на выполнение гарантийных обязательств и сервисное обслуживание.

7. *На этапе утилизации.* Моделирование изделия при утилизации проводится для определения возможных методов и средств выполнения этой задачи, аргументируется также повторное использование деталей или узлов в зависимости от определенного срока их службы (например, если количество времени до разрушения детали значительно превышает срок службы изделия в целом).

## **2. Перспективы развития аналитических средств поддержки жизненного цикла изделий**

Как следует из приведенного выше материала и примеров использования аналитических средств на различных этапах ЖЦИ, информационные технологии прочно обосновались в практике машиностроительных предприятий (включая гидроприводостроение). Сложившаяся тенденция в развитии интерфейсов программных систем (ленточных, контекстно-зависимых) уже сейчас обеспечивает смену функционала системы при том же информационном содержании активного экрана. При сохранении производителями программных продуктов данной тенденции следует рассчитывать на создание единого интерфейса, меняя содержимое которого, можно будет «пролистать» ЖЦИ от замысла дизайнера до конструктивной проработки, от технологической подготовки производства к изготовлению и эксплуатации изделия – до его утилизации. Развитие вычислительной техники и прикладного программного обеспечения позволит анализировать изделия более подробно, проводить исследования взаимовлияния факторов конструкторской проработки, технологических ограничений, производственных возможностей и экономических факторов.

Кроме того, развитие математического аппарата, программных комплексов и средств вычислительной техники позволит более широко использовать мультидисциплинарный подход к аналитическому сопровождению ЖЦИ [6]. Данная информационная технология даст возможность более реалистично описывать модели изделия, быстрее получать результат (без итераций между однодисциплинарными моделями) и, соответственно, иметь на выходе увеличение качества и надежности изделий гидроприводов и уменьшение количества натуральных экспериментов и брака.

### **Заключение**

Концепция ЖЦИ на практике применяется на большинстве машиностроительных предприятий. Использование единых стандартов при формировании моделей позволяет использовать их на всех этапах ЖЦИ, а также на других (смежных) предприятиях.

Интеграция систем управления ЖЦИ в бизнес-процессы предприятия обеспечивает его стратегические преимущества, включая повышение эффективности и темпов производства, уменьшение числа ошибок, сокращение сроков выхода изделий на рынок.

Использование единых форматов данных и принципов формирования семейства моделей изделия обеспечивает передачу информации без потерь между этапами эскизного и технического проектирования, численного моделирования и изготовления, что значительно сокращает сроки подготовки производства. Применение единого информационного пространства и современного мультидисциплинарного программного обеспечения позволяет проводить аналитические исследования на всех стадиях ЖЦИ, уменьшать количество фальш-итерационных процессов и реальных опытных образцов.

В данной статье приведен общий подход к аналитическому сопровождению жизненного цикла гидроприводов, поскольку в настоящее время подобный комплексный подход реализован на предприятии лишь частично. Поэтому в ближайшее время предстоит интенсификация работ в данном направлении, результаты которых будут представлены в следующих статьях автора.

### Список литературы

1. Стратегия развития отрасли информационных технологий в Российской Федерации на 2014–2020 годы и на перспективу до 2025 года: утв. распоряжением Правительства РФ от 1 ноября 2013 г. № 2036-р. – М., 2013. – 51 с.
2. Система разработки и постановки продукции на производство. Термины и определения : Р 50-605-80-93. – Утв. 9.07.1993. – М. : Изд-во стандартов, 1993. – 45 с.
3. Информационная технология. Системная инженерия. Процессы жизненного цикла систем : ГОСТ Р ИСО/МЭК 15288–2005. – Введ. 29.12.2005. – М. : Изд-во стандартов, 2005. – 65 с.
4. Норенков, И.П. Автоматизированное проектирование : учебник. Сер. Информатика в техническом университете / И.П. Норенков. – М. : Изд-во МГТУ им. Н.Э. Баумана, 2000. – 188 с.
5. Баккер, М. Поддержка моделирования работы предприятий авиационно-космической и оборонной отраслей / М. Баккер // CAD/CAM/CAE Observer. – 2014. – № 6(90). – С. 24–29.
6. Пузанов, А.В. Опыт использования современного мультифизического ПО в разработке электрогидроприводов / А.В. Пузанов // САПР и графика. – 2008. – № 4. – С. 75–79.
7. Криворучко, Д.В. Моделирование процессов обработки материалов резанием / Д.В. Криворучко, М. Г. Сторчак, В.А. Залого // Оборудование и инструмент для профессионалов. – 2009. – № 4(115). – С. 46–48.
8. Пестрецов, С.И. Компьютерное моделирование и оптимизация процессов резания : учеб. пособие / С.И. Пестрецов. – Тамбов : Изд-во Тамбовского гос. техн. ун-та, 2009. – 104 с.
9. Комаров, В.Н. Математическое моделирование. Оптимальное управление / В.Н. Комаров, А.В. Грезина, С.А. Артемьева // Вестник Нижегородского университета им. Н.И. Лобачевского. – 2011. – № 3(2). – С. 87–91.
10. Даршт, Я.А. Методы и модели автоматизированного анализа и синтеза элементов гидропривода : автореф. дис. ... д-ра техн. наук : 05.13.12 / Я.А. Даршт ; Владимирский гос. ун-т. – Ковров, 2005. – 43 с.
11. Воронов, С.А. Методы расчета и проектирования аксиально-поршневых гидромашин силовых регулируемых гидроприводов : автореф. дис. ... д-ра техн. наук : 05.02.02 / С.А. Воронов ; Ковровская гос. технол. акад. – Ковров, 2008. – 36 с.
12. Пузанов, А.В. Автоматизация конструкторско-технологической подготовки производства в ОАО «СКБ ПА» / А.В. Пузанов, К.Е. Куванов, А.Н. Часовских // САПР и графика. – 2009. – № 11. – С. 25–28.
13. Пузанов, А.В. Расчет усталости в Autodesk Simulation Multiphysics / А.В. Пузанов // Autodesk community magazine. – 2013. – № 1(4). – С. 48–52.

Поступила 11.01.2016

*ОАО «Специальное конструкторское бюро  
приборостроения и автоматики»,  
Ковров, ул. Крупской, 55  
e-mail: avp@oao-skbpa.ru*

**A.V. Puzanov**

**INFORMATION SUPPORT OF HYDRODRIVES LIFECYCLE**

The paper discusses the ways of increasing the efficiency of production of hydraulic drives by using the information support of product life cycle. The article describes examples of the use of software systems for analyzes at different stages of the life cycle of hydraulic drives.

## ОБРАБОТКА СИГНАЛОВ, ИЗОБРАЖЕНИЙ И РЕЧИ

УДК 004.932.2

В.А. Левчук, В.А. Ковалев

## МЕТОДИКА КОЛИЧЕСТВЕННОГО ОПИСАНИЯ БИМЕДИЦИНСКИХ ИЗОБРАЖЕНИЙ НА ОСНОВЕ СЛОВАРЕЙ СУПЕРПИКСЕЛОВ

*Предлагается методика количественного описания биомедицинских изображений, основанная на разбиении целевого изображения на суперпикселы и их сопоставлении с ранее подготовленным словарем суперпикселов, характерных для изображений анализируемого типа. Методика протестирована на задачах распознавания биомедицинских изображений трех типов (КТ-снимков легкого, гистологических изображений образцов тканей яичников и тканей щитовидной железы). Экспериментально показывается, что предлагаемая методика обеспечивает результаты, сравнимые по качеству распознавания с традиционными методами описания структуры изображений либо превосходящие их.*

**Введение**

Одним из важнейших этапов в задачах обработки и анализа биомедицинских изображений является их количественное описание. От выбора нужной методики вычисления количественных параметров биомедицинских изображений может сильно зависеть качество работы алгоритмов, используемых при распознавании изображений, поиске в базах данных по содержанию [1], в системах компьютеризированной диагностики [2]. Недавние исследования говорят о первостепенной важности этапа количественного описания изображений для более точного их распознавания и понимания [3]. Одним из направлений развития методик количественного описания изображений является переход от прямолинейного использования достаточно простых локальных характеристик (яркостей пикселей, градиентов) к описанию на основе более информативных морфологических и геометрических базовых структур (визуальных слов [4], патчей [5] и др.).

В последнее время активно развивается и получает все более широкое применение концепция представления изображения при помощи так называемых суперпикселов (в случае трехмерных изображений – супервокселов). Суперпикселы представляют собой небольшие относительно однородные участки, целиком покрывающие все исходное изображение. На сегодняшний день суперпиксельное представление используется в различных задачах, таких как локализация объектов [6], скелетизация структур [7] и распознавание сцен [8]. Однако благодаря своей способности адаптивно разбивать изображение на участки, границы которых часто совпадают с очертаниями объектов на изображении, наибольшее применение суперпикселы находят в задачах сегментации изображений различных типов – от уличных сцен до медицинских снимков [9–11].

Целью данной работы является разработка методики описания биомедицинских изображений с использованием их суперпиксельного представления. Предлагаемая методика основана на применении словарей суперпикселов, характерных для исследуемого типа изображений. Отметим, что в настоящее время проблема описания изображений при помощи суперпиксельного подхода крайне мало изучена. Так, в [12] для описания изображений использовался подход, основанный на вычислении яркостных различий соседних суперпикселов, расположенных под различными углами, что слабо раскрывает потенциал использования такого представления изображений для описания их содержания.

**1. Используемые материалы**

В настоящей работе для оценки эффективности предлагаемой методики описания изображений использовались три набора данных.

Участки КТ-снимков легкого. Типичные КТ-изображения легкого обычно представляют собой трехмерные цифровые снимки, состоящие примерно из 60–300 двумерных слоев размером 512×512 пикселей. Из исходных трехмерных КТ-снимков 195 пациентов, больных туберкулезом легких, были вырезаны в общей сложности 270 двумерных участков размером 128×128 пикселей (рис. 1, а, б) с яркостным окном от –1000 до +300 HU (Hounsfield Units). Из их числа 92 изображения соответствовали участкам легкого, пораженным туберкулезом, а остальные 178 – участкам без видимых признаков заболевания. Задача автоматического распознавания пораженных участков легкого используется при выделении наиболее информативных слоев на КТ-снимках, а также при подсвечивании на изображениях признаков заболевания.

Наборы гистологических изображений яичников (рис. 1, в, г) и щитовидной железы (рис. 1, д, е). На каждый орган имелось по 200 изображений, 100 из которых были снимками участков здоровой ткани, а остальные 100 соответствовали тканям, пораженным опухолью. В рассматриваемом эксперименте все гистологические изображения были приведены к размеру 256×256 пикселей и преобразованы в полутоновые. В работе [13] отмечается, что на подобных гистологических изображениях более сильно проявляют себя структурные признаки, нежели цветовые. Предлагаемая в работе методика описания изображений тестировалась на задаче распознавания гистологических снимков на классы, представленные здоровой и пораженной тканью.

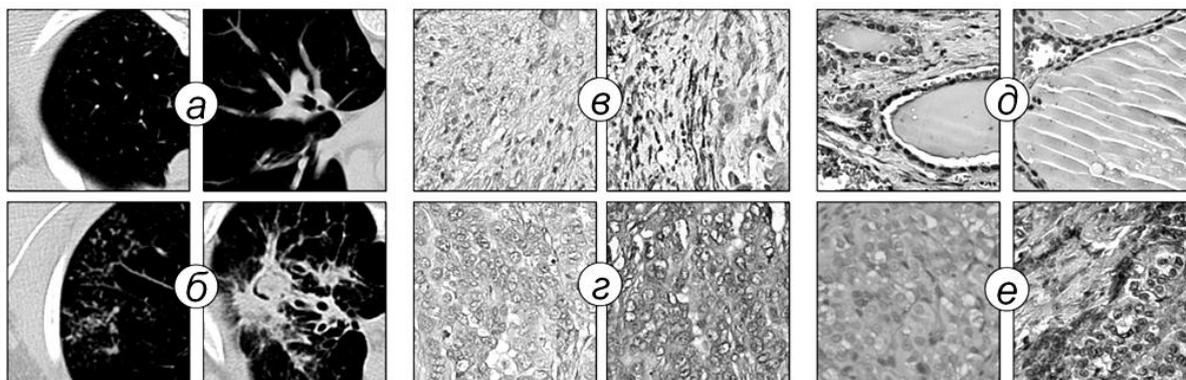


Рис. 1. Примеры исходных изображений: а) КТ-изображений здоровых тканей; б) КТ-изображений пораженных тканей; в) здоровых тканей яичников; г) пораженных тканей яичников; д) здоровых тканей щитовидной железы; е) пораженных тканей щитовидной железы

## 2. Общая схема методики

Предлагаемая методика описания биомедицинских изображений предполагает выполнение двух основных этапов: создание словаря суперпикселей и описание изображений при помощи полученного словаря.

Используемый словарь суперпикселей должен в явном либо неявном виде содержать информацию о наиболее типичных суперпикселях, встречающихся на изображениях рассматриваемого типа. При решении задачи распознавания словари представлялись набором количественных характеристик наиболее типичных суперпикселей. Этап создания словаря суперпикселей включал следующие шаги:

- выбор некоторого количества представительных изображений заданного типа;
- разбиение выбранных изображений на суперпиксели;
- выделение характеристик полученного множества суперпикселей;
- разбиение множества всех суперпикселей на заданное количество  $N$  классов в пространстве признаков (кластеризацию);
- нахождение центроидов классов;
- создание словаря в виде матрицы чисел, каждая из  $N$  строк которой представляет собой вектор характеристик центроида соответствующего класса суперпикселей.

Описание изображений при помощи ранее полученных словарей производилось на основе подсчета на изображениях количества суперпикселей, принадлежащих к  $N$  различным классам. Таким образом, процедура описания изображения включала следующие шаги:

- разбиение целевого изображения на суперпиксели;
- выделение характеристик каждого суперпикселя;
- вычисление расстояний между векторами характеристик суперпикселей целевого изображения и элементами словаря (векторами характеристик центроидов);
- отнесение каждого суперпикселя к определенному классу в соответствии с минимальным расстоянием в пространстве признаков;
- построение гистограммы распределения суперпикселей изображения по классам.

Полученная в итоге гистограмма служит дескриптором целевого изображения. Отметим, что существуют и другие, более сложные варианты построения конечного дескриптора, например построение матрицы встречаемости классов соседних суперпикселей [10] и др. В данной работе авторы ограничились рассмотрением лишь предложенного метода, основанного на гистограмме классов суперпикселей.

### 3. Количественные характеристики суперпикселей

В литературе встречаются различные способы описания отдельных суперпикселей для решения разного рода задач. Так, в [14] для задачи сегментации сцен использовалось в общей сложности 1708 характеристик, описывающих цветовые, текстурные свойства суперпикселей, свойства формы, а также их местоположение. Поскольку настоящая работа посвящена анализу полутоновых изображений, а количество пикселей в суперпикселе в рассматриваемом случае было невелико и обычно составляло около 200–300 штук, было решено отказаться от использования сложных текстурных дескрипторов и статистик высоких порядков. Информация о местоположении отдельных суперпикселей на изображении также не использовалась.

Для задачи количественного описания полутоновых изображений были использованы шесть характеристик, описывающих содержание и форму каждого отдельно взятого суперпикселя: средняя яркость внутренних пикселей, стандартное отклонение яркостей внутренних пикселей, энтропия яркостей внутренних пикселей, средняя величина градиента внутри суперпикселя, компактность, квадратность.

Энтропия яркостей вычислялась стандартным образом при помощи гистограммы, различающей восемь градаций яркости пикселей. Значение каждого бина ненормализованной гистограммы равно количеству пикселей соответствующей яркости внутри суперпикселя. Количество бинов гистограммы принято равным восьми, поскольку в случае больших значений гистограммы получаются разреженными ввиду ограниченного количества пикселей изображения в суперпикселе. Так, например, при 256 бинах гистограмма может состоять преимущественно из нулей и единиц. Для вычисления величины градиента использовался оператор Собеля.

Величина компактности, характеризующая форму суперпикселя, вычислялась по формуле

$$c = \frac{\sqrt{S}}{P},$$

где  $P$  – количество граничных пикселей (периметр),  $S$  – общее количество пикселей (площадь) суперпикселя. Таким образом, суперпиксели более сложной формы имели меньшие значения компактности.

Еще одним параметром, характеризующим форму суперпикселя, была так называемая квадратность, которая выражает степень близости формы к квадрату. Использование такого количественного параметра целесообразно, поскольку в применяемой реализации алгоритма построения суперпикселей [15] на больших однородных участках изображения последние имели форму, близкую к квадрату. Степень квадратности вычислялась как отношение площади суперпикселя к площади его ограничивающего прямоугольника (axis-aligned bounding box). Вычисляемая таким образом характеристика дополнительно показывает степень искажения формы суперпикселя как отклик на содержание исходного изображения.

#### 4. Словари суперпикселов

В данной работе словари суперпикселов были сгенерированы для трех типов данных: участков КТ-снимков, гистологических изображений яичников и гистологических изображений щитовидной железы. Используемый в работе алгоритм построения суперпикселов [15] имеет два управляющих параметра: начальный размер суперпикселов  $Sz$  и параметр регуляризации  $Reg$  (определяет способность суперпикселов к деформации). От параметра  $Sz$  зависит итоговое количество суперпикселов на изображении. Оно определяется как отношение площади исходного изображения к квадрату величины  $Sz$ . Примеры работы алгоритма на двух тестовых изображениях приведены на рис. 2.

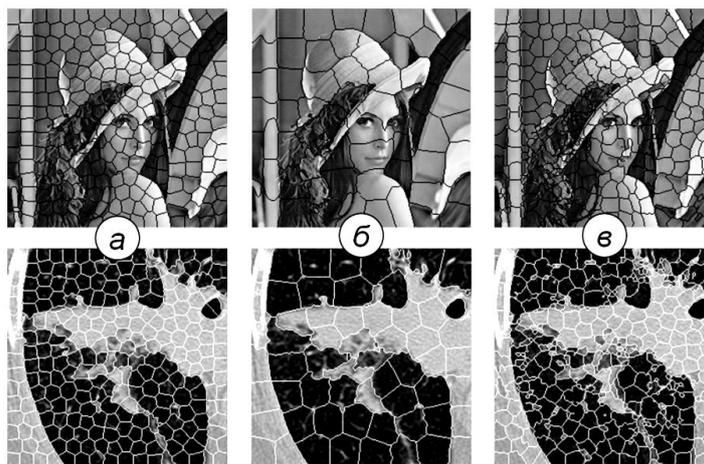


Рис. 2. Примеры работы алгоритма построения суперпикселов с различными параметрами:  
 а)  $Sz = 16$ ,  $Reg = 0,1$ ; б)  $Sz = 32$ ,  $Reg = 0,1$ ; в)  $Sz = 16$ ,  $Reg = 0,01$

Словари суперпикселов генерировались отдельно для каждой рассматриваемой комбинации параметров  $Sz$  и  $Reg$ . Для всех трех наборов исходных данных из общего числа суперпикселов на всех изображениях случайным образом выбирались по 10 000 суперпикселов. Кластеризация выбранного количества суперпикселов в пространстве признаков производилась при помощи метода  $k$ -средних с различными значениями количества кластеров  $N = 16, 32, 64, 128, 256$ . Полученные в результате словари представляли собой матрицы чисел размером  $N \times 6$ , в которых число столбцов соответствует количеству характеристик суперпикселов, а число строк – количеству кластеров  $N$  (размеру словаря суперпикселов).

На рис. 3 показаны характерные примеры суперпикселов, полученных на участках КТ-изображений с параметрами  $Sz = 16$  и  $Reg = 0,1$ . Каждый столбец соответствует одному классу суперпикселов.

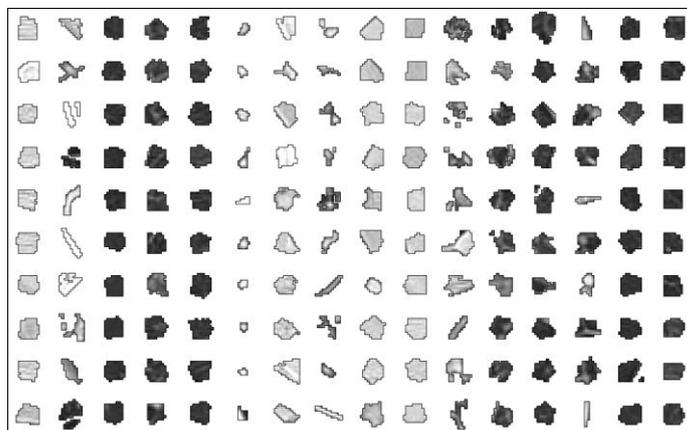


Рис. 3. Примеры суперпикселов участков КТ-снимков, принадлежащих 16 классам

## 5. Тестирование методики на задаче распознавания изображений

Оценка работоспособности методики описания изображений, предлагаемой в данной работе, проводилась на задаче распознавания изображений. Поскольку целью данной работы являлось не достижение максимальной точности распознавания, а сравнение эффективности работы дескрипторов, в качестве классификатора был использован метод  $K$  ближайших соседей. Такой подход позволяет упростить задачу оценки эффективности работы дескриптора за счет отказа от использования сложных классификаторов. Для вычисления расстояний между дескрипторами изображений использовалась метрика городских кварталов  $L1$ , представляющая собой нормированную сумму разностей элементов дескрипторов. Как показывают исследования, данная метрика хорошо подходит к дескрипторам гистограммного типа [16]. Для оценки ошибки определения точности распознавания использовалась статистическая процедура Bootstrap [17]. Во время этой процедуры исходная выборка данных многократно случайным образом разбивается на подвыборки, на которых проводится оценка качества распознавания. Под подвыборкой в данном случае подразумевается подмножество исходного множества изображений без повторов. В рассматриваемом эксперименте размер подвыборки составлял 80 % от размера исходной выборки, процедура разбиения повторялась 100 раз. Итоговая точность распознавания вычислялась как среднее значение точностей, полученных на подвыборках, а ошибка оценки точности – как их среднеквадратичное отклонение.

Оценка проводилась для следующих комбинаций управляющих параметров алгоритма построения суперпикселов:  $Sz = 16, Reg = 0,1$ ;  $Sz = 32, Reg = 0,1$ ;  $Sz = 8, Reg = 0,1$ ;  $Sz = 16, Reg = 0,3$ ;  $Sz = 16, Reg = 0,03$ ;  $Sz = 16, Reg = 0,01$ . Размер словаря суперпикселов  $N$  в данном эксперименте принимал значения 16, 32, 64, 128 и 256. Количество ближайших соседей  $K$  было выбрано равным пяти (подобрано эмпирически). В качестве основного дескриптора использовались гистограммы распределения суперпикселов по заданному количеству классов. Для сравнения с существующими методами аналогичный эксперимент был проведен с применением локальных бинарных шаблонов (LBP) [18] и матриц совместной встречаемости (co-occurrence matrices) [13] в качестве дескрипторов изображений. Для метода локальных бинарных шаблонов использовались маски радиусом 1, 3, 5 и 9 пикселов. Используемые в работе матрицы встречаемости различали восемь градаций интенсивности и учитывали пары пикселов на расстояниях 1, 3 и 5.

В табл. 1–4 представлены результаты оценки точности распознавания изображений, полученные с использованием различных дескрипторов. Из рассмотренных комбинаций управляющих параметров предлагаемого метода описания изображений наилучшие результаты распознавания были достигнуты для КТ-изображений при  $Sz = 16, Reg = 0,3$ ; для изображений яичников – при  $Sz = 8, Reg = 0,1$ ; для изображений щитовидной железы – при  $Sz = 16, Reg = 0,03$ .

Таблица 1

Точность распознавания участков КТ-изображений

Управляющие параметры		Размер словаря суперпикселов				
$Sz$	$Reg$	16	32	64	128	256
16	0,1	0,892	0,888	0,885	0,893	0,871
32	0,1	0,89	0,874	0,864	0,852	0,874
8	0,1	0,87	0,888	0,882	0,891	0,891
16	0,3	0,894	<b>0,91</b>	0,885	0,88	0,875
16	0,03	0,878	0,884	0,88	0,88	0,867
16	0,01	0,872	0,867	0,868	0,881	0,853

Таблица 2

Точность распознавания гистологических изображений яичников

Управляющие параметры		Размер словаря суперпикселей				
$Sz$	$Reg$	16	32	64	128	256
16	0,1	0,832	0,839	0,882	0,876	0,871
32	0,1	0,853	0,856	0,854	0,864	0,847
8	0,1	0,864	0,857	<b>0,894</b>	0,875	0,871
16	0,3	0,834	0,853	0,84	0,865	0,855
16	0,03	0,864	0,87	0,851	0,865	0,87
16	0,01	0,876	0,868	0,852	0,881	0,85

Таблица 3

Точность распознавания гистологических изображений щитовидной железы

Управляющие параметры		Размер словаря суперпикселей				
$Sz$	$Reg$	16	32	64	128	256
16	0,1	0,843	0,87	0,885	0,895	0,897
32	0,1	0,899	0,874	0,893	0,876	0,889
8	0,1	0,878	0,898	0,897	0,896	0,913
16	0,3	0,857	0,881	0,873	0,884	0,892
16	0,03	0,903	0,898	0,895	0,906	<b>0,921</b>
16	0,01	0,895	0,905	0,902	0,907	0,906

Таблица 4

Точность распознавания традиционными методами

Набор данных	Метод распознавания	
	Матрицы совместной встречаемости	Локальные бинарные шаблоны
Участки КТ-изображений	0,887	0,849
Гистология (яичники)	0,868	0,850
Гистология (щитовидная железа)	0,894	0,837

На графиках (рис. 4) сравнивается точность распознавания для различных типов дескрипторов. Точность распознавания с использованием дескриптора, основанного на суперпикселях, приведена для значений параметров  $Sz$  и  $Reg$ , дающих наилучшую точность распознавания. Как видно из графиков, предлагаемый метод описания изображений позволяет повысить точность распознавания по сравнению с использованием традиционных матриц совместной встречаемости и локальных бинарных шаблонов. В частности, увеличение точности распознавания по сравнению с использованием матриц совместной встречаемости достигает 2,3 п. п. в случае участков КТ-изображений, 2,6 п. п. в случае гистологических изображений яичников и 2,7 п. п. в случае гистологических изображений щитовидной железы.

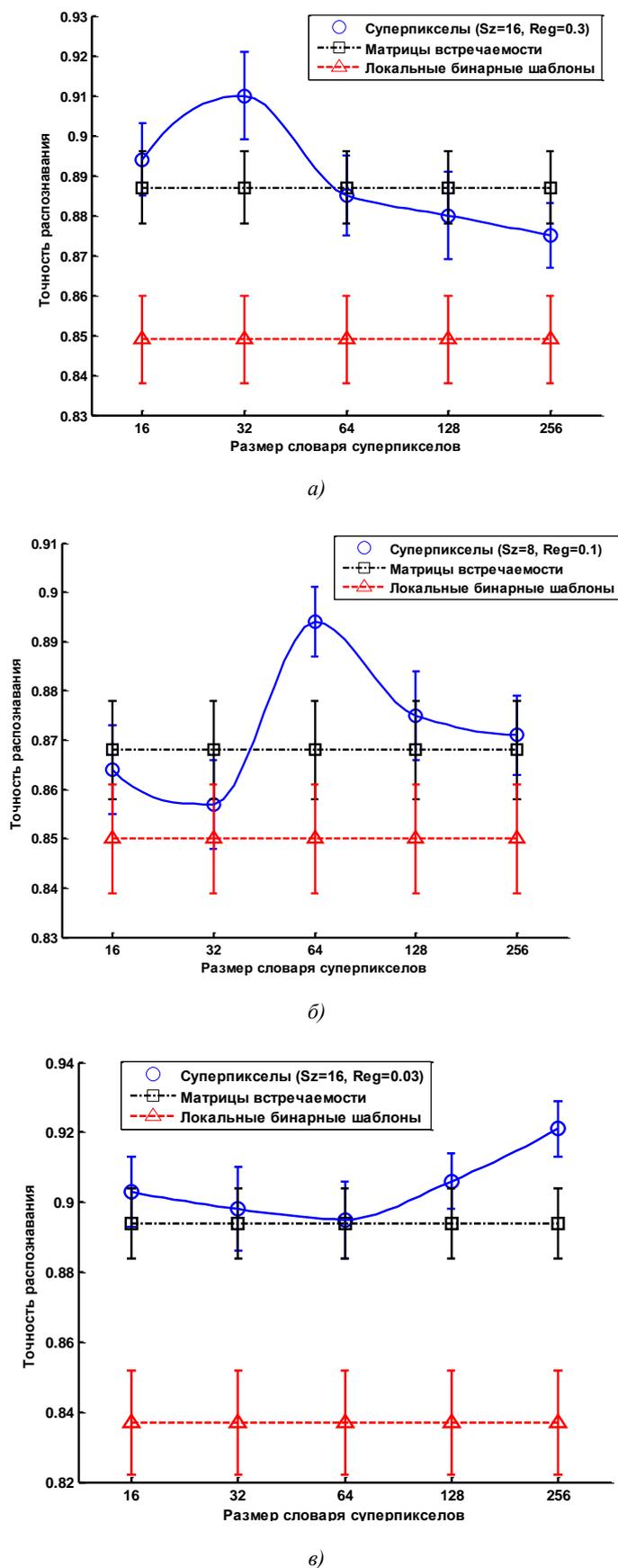


Рис. 4. Оценка точности распознавания на разных типах данных: а) участки КТ-изображений; б) гистологические изображения яичников; в) гистологические изображения щитовидной железы

### **Заключение**

Предложенный метод описания полутоновых биомедицинских изображений, основанный на использовании словарей суперпикселов, которые характерны для изображений анализируемого типа, предполагает разбиение целевого изображения на суперпикселы, отнесение полученных суперпикселов к классам в соответствии с ранее подготовленным словарем и подсчет количества суперпикселов каждого класса. На примере задач распознавания участков КТ-снимков и гистологических изображений в работе было показано, что применение предлагаемой методики описания изображений может повысить точность решения задачи распознавания по сравнению с использованием таких методов, как матрицы совместной встречаемости и локальные бинарные шаблоны.

Описываемый метод может быть достаточно просто модифицирован и адаптирован к цветным либо трехмерным изображениям. Отметим также, что сама концепция описания изображений при помощи разбиения на суперпикселы может быть реализована и без применения словарей суперпикселов. К примеру, дескриптор изображения может быть построен напрямую из гистограмм распределения характеристик суперпикселов целевого изображения. Разработанная методика требует дальнейшего развития и тестирования на других задачах с использованием различных типов изображений.

### **Список литературы**

1. Kovalev, V. Color Co-occurrence Descriptors for Querying-by-Example / V. Kovalev, S. Volmer // Proc. of the 1998 Conf. on MultiMedia Modeling. – Switzerland, 1998. – P. 32–38.
2. Content-based image retrieval as a method for melanoma diagnosis / V. Liauchuk [et al.] // Proc. of the 29th Intern. Congress on Computer Assisted Radiology and Surgery (CARS–2015). – Spain, 2015. – Vol. 10. – P. 291–292.
3. Pinto, N. A high-throughput screening approach to discovering good forms of biologically inspired visual representation / N. Pinto, D. Doukhan, D. Cox // PLoS Computational Biology. – 2009. – Vol. 5(11). – P. 1–12.
4. Evaluating Bag-of-Visual-Words Representations in Scene Classification / J. Yang [et al.] // Proc. of the Intern. Workshop on Multimedia Information Retrieval. – Germany, Augsburg, 2007. – P. 197–206.
5. Quantized Image Patches Co-occurrence Matrix: A New Statistical Approach for Texture Classification using Image Patch Exemplars / Zh. Liu [et al.] // Proc. of the 3th Intern. Conf. on Digital Image Processing (ICDIP 2011). – China, Chengdu, 2011. – Vol. 8009. – P. 56–64.
6. Fulkerson, B. Class segmentation and object localization with superpixel neighborhoods / B. Fulkerson, A. Vedaldi, S. Soatto // Proc. of the 12th Intern. Conf. on Computer Vision. – Kyoto, Japan. – 2009. – P. 8.
7. Levinshtein, A. Multiscale symmetric part detection and grouping / A. Levinshtein, C. Sminchisescu, S. Dickinson // Proc. of the 12th Intern. Conf. on Computer Vision. – Kyoto, Japan. – 2009. – Vol. 104. – P. 117–134.
8. Li, X. An Object Co-occurrence Assisted Hierarchical Model for Scene Understanding / X. Li, Y. Guo // Proc. of the British Machine Vision Conference. – Surrey, 2012. – P. 1–11.
9. Multi-scale feature learning on pixels and super-pixels for seminal vesicles MRI segmentation / Q. Gao [et al.] // SPIE 9034, Medical Imaging 2014: Image Proc. – 2014. – Vol. 9034. – P. 903407–903407-6.
10. Micusik, B. Semantic segmentation of street scenes by superpixel co-occurrence and 3D geometry / B. Micusik, J. Kosecka // IEEE Workshop on Video-Oriented Object and Event Classification (VOEC). – Japan, 2009. – P. 625–632.
11. He, X. Learning and incorporating top-down cues in image segmentation / X. He, R. Zemel, D. Ray // Proc. of the 9th European Conf. on Computer Vision. – Austria, 2006. – P. 338–351.
12. Sicre, R. SuperPixel based angular differences as a mid-level image descriptor / R. Sicre, E. Tasli, T. Gevers // Proc. of the 22th Intern. Conf. on Pattern Recognition (ICPR). – Sweden, 2014. – P. 3732–3737.

13. Content-based image retrieval of histology images / A. Dmitruk [et al.] // Proc. of Swedish Symposium on Image Analysis (SSBA-2010). – Uppsala, Sweden, 2010. – P. 91–92.
14. Tighe, J. SuperParsing: Scalable nonparametric image parsing with superpixels / J. Tighe, S. Lazebnik // ECCV'10 Proc. of the 11th European Conf. on Computer Vision. – Germany, Heidelberg, 2010. – P. 352–365.
15. SLIC Superpixels Compared to State-of-the-art Superpixel Methods / R. Achanta [et al.] // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 2012. – Vol. 34(11). – P. 2274–2282.
16. Ковалев, В.А. Влияние мер близости в пространстве признаков на качество поиска медицинских изображений по содержанию / В.А. Ковалев, А.А. Дмитрук // Информатика. – 2011. – № 30. – С. 5–11.
17. Efron, B. Bootstrap Methods: Another Look at the Jackknife / B. Efron // Annals of Statistics. – 1979. – Vol. 1. – P. 1–26.
18. Matas, T. Extended set of local binary patterns for rapid object detection / T. Matas, J. Matas // Proc. of the Computer Vision Winter Workshop. – Czech Republic, Nové Hrad, 2010. – Vol. 2010. – P. 37–43.

Поступила 09.11.2015

*Объединенный институт проблем  
информатики НАН Беларуси,  
Минск, Сурганова, 6  
e-mail: vitali.liauchuk@gmail.com*

**V.A. Liauchuk, V.A. Kovalev**

## **A METHOD FOR QUANTITATIVE DESCRIPTION OF BIOMEDICAL IMAGES BASED ON SUPERPIXEL DICTIONARIES**

With this study, a method for quantitative description of biomedical images based on splitting the target image into superpixels followed by categorization using precalculated superpixel dictionaries is proposed. The method has been tested on the tasks of recognition of biomedical images of three types: lung CT images, histology images of ovary and thyroid tissues. The results of the experiments performed suggest that the method proposed may provide recognition performance comparable or better than when using conventional methods of texture description.

УДК 620.179.15

М.А. Мирзаванд

## ПАРАЛЛЕЛЬНАЯ ИТЕРАЦИОННАЯ РЕКОНСТРУКЦИЯ ФАНТОМА САТРИАН ПО ЭКСПЕРИМЕНТАЛЬНЫМ ДАННЫМ

*Рассматриваются принципы создания быстрых параллельных итерационных алгоритмов, основанных на использовании графических ускорителей и библиотеки OpenGL. Предложенные подходы обеспечивают одновременную минимизацию невязки искомого решения и тотальной вариации реконструируемого трехмерного изображения. При этом число необходимых исходных данных (конических рентгеновских проекций), а значит, и лучевая нагрузка на пациента при сохранении необходимого контраста и пространственного разрешения трехмерного изображения могут быть уменьшены в несколько раз. Предложенный эвристический итерационный алгоритм может быть использован как альтернатива известному трехмерному алгоритму Фельдкамп.*

### Введение

Как в промышленности, так и в медицине широкое распространение получила компьютерная томография с использованием конического пучка рентгеновских лучей *СВСТ* (*Cone Beam Computed Tomography*). Так, *СВСТ* используется в модулированной по интенсивности лучевой терапии *IMRT* (*Intensity-Modulated Radiation Therapy*), которая направлена на максимальную защиту соседних с опухолью здоровых тканей, при этом сама опухоль получает максимальную необходимую дозу. Лучевая терапия обычно проводится на линейном ускорителе последнего поколения Электа Синержи (*Elekta Synergy Platform*), работающем по принципу *IMRT*.

Самыми распространенными на сегодняшний день методами реконструкции по коническим проекциям являются методы, основанные на использовании алгоритма Фельдкамп. В линейном ускорителе Электа Синержи имеется киловольтная рентгеновская трубка, с помощью которой перед сеансом лучевой терапии производится *СВСТ* пациента. Стандартным способом реконструкции является алгоритм Фельдкамп. Вследствие наметившейся в последние годы тенденции к значительному уменьшению рентгеновской дозы для пациентов, подвергаемых томографическому обследованию, возникает необходимость уменьшения как числа снимаемых проекций, так и мощности анодного тока рентгеновской трубки, при которой проводится съемка. При компьютерной томографии пациента, как правило, используются порядка 300–700 рентгеновских проекций с лучевой нагрузкой 0,4 мАс на одну проекцию. С уменьшением рентгеновской дозы до 0,1 мАс на одну проекцию или числа снимаемых проекций до 30–100 алгоритм Фельдкамп реконструирует сильно искаженные изображения пациентов, которые непригодны для дальнейшего медицинского использования.

Исправить эту ситуацию можно с помощью итерационных томографических алгоритмов с использованием в качестве регуляризирующего функционала тотальной вариации (*total variation, TV*). Метод *TV* обеспечил новые возможности итерационной томографии, позволяющие успешно реконструировать изображения с помощью минимизации полунормы *TV* даже при ограниченном числе зашумленных проекций. Использование минимизации *TV* реконструируемого изображения позволяет устранить артефакты, вызванные зашумлением рентгеновских проекций, которое возникает вследствие малой мощности тока рентгеновской трубки, а также обеспечивает практически полное устранение полосовых артефактов, вызванных уменьшением числа проекций. Задача состоит в том, чтобы разработать итерационные алгоритмы томографической реконструкции с использованием минимизации *TV*, позволяющие по ограниченному числу проекций (30–100) и рентгеновской дозе 0,1–0,2 мАс на проекцию обеспечить практически такое же качество реконструируемого изображения, как и алгоритм Фельдкамп, для 300–700 проекций с лучевой нагрузкой на одну проекцию 0,4 мАс.

Между тем стандартные итерационные методы томографии затрачивают на реконструкцию трехмерного изображения пациента неприемлемо большое время. Использование графических ускорителей позволяет разработать для трехмерной томографической реконструкции с одновременной минимизацией  $TV$  сверхбыстрые параллельные алгоритмы, которые решают эту задачу за временной промежуток от нескольких секунд до одной-двух минут.

### 1. Эвристический метод статистической итерационной реконструкции

В настоящей работе рассмотрена возможность улучшения качества стандартной томографической реконструкции путем использования вместо нее параллельной итерационной реконструкции с применением библиотеки *OpenGL* и регуляризацией. Кроме того, было оценено качество реконструкции по малому числу проекционных видов. Итерационные способы томографии основаны на решении системы линейных алгебраических уравнений

$$Ax = p, \quad (1)$$

где  $A = (A_{ij})$  – проекционная матрица;  $\mathbf{x} = (x_1, x_2, \dots, x_J)$  – вектор изображения;  $\mathbf{p} = (p_1, \dots, p_I)$  – вектор проекций. Таким образом, решение основной задачи восстановления изображения по заданному набору проекций  $p_i$  сводится к решению системы линейных алгебраических уравнений вида (1), при этом вектор  $\mathbf{p}$  заведомо задан с некоторой погрешностью. Определим

$$A_{i+} = \sum_j A_{ij} \quad (2)$$

для  $i=1, \dots, I$ ,

$$A_{+j} = \sum_i A_{ij} \quad (3)$$

для  $j=1, \dots, J$ .

Тогда формула для реконструкции методом *SART* примет вид

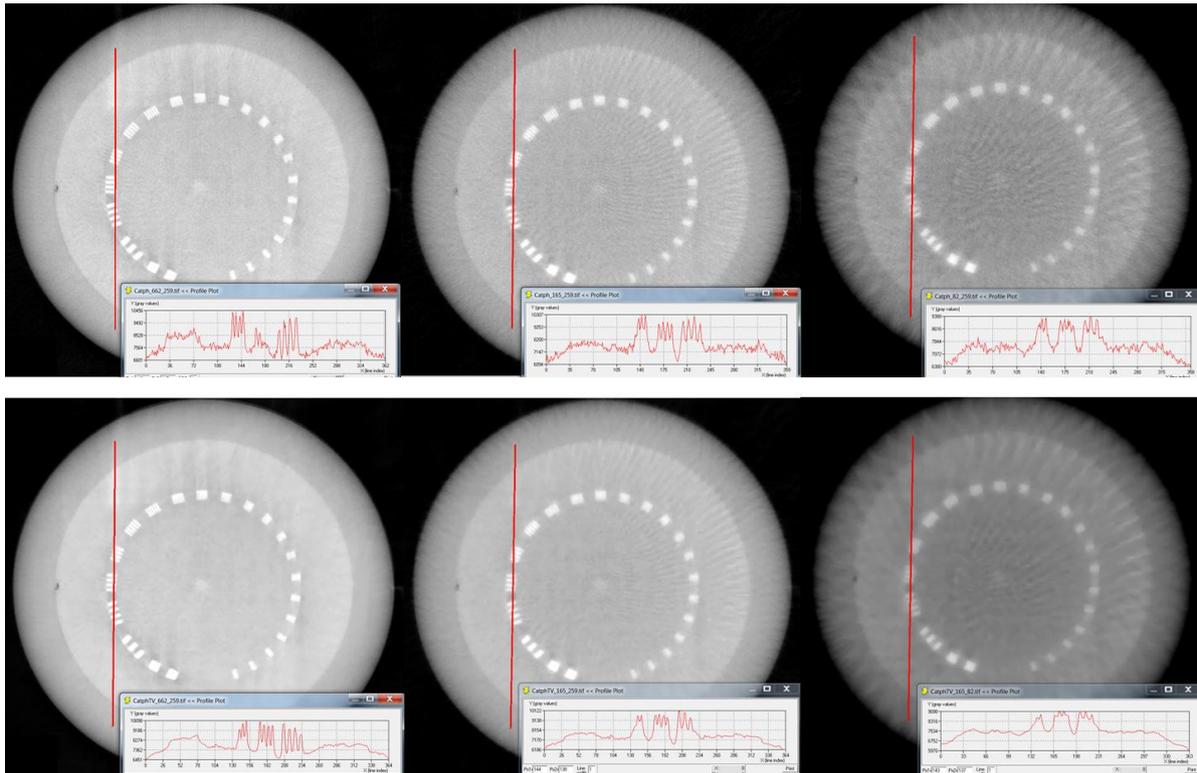
$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \lambda^{(k)} / A_{+j} \sum_i A_{ij} / A_{i+} (p_i - (A^i, \mathbf{x}^{(k)})), \quad (4)$$

где параметры релаксации  $\lambda^{(k)}$  представляют собой последовательность вещественных чисел ( $0 < \lambda^{(k)} < 2$ ). В работе исследуется эвристический итерационный алгоритм [1], по теоретическим соображениям близкий к алгоритму, опубликованному в работе [2], но отличающийся прежде всего тем, что в нем используется коррекция текущего приближения с помощью минимизации невязки для лучевых интегралов, а не для интенсивностей. Предлагаемый алгоритм реконструкции можно записать следующим образом:

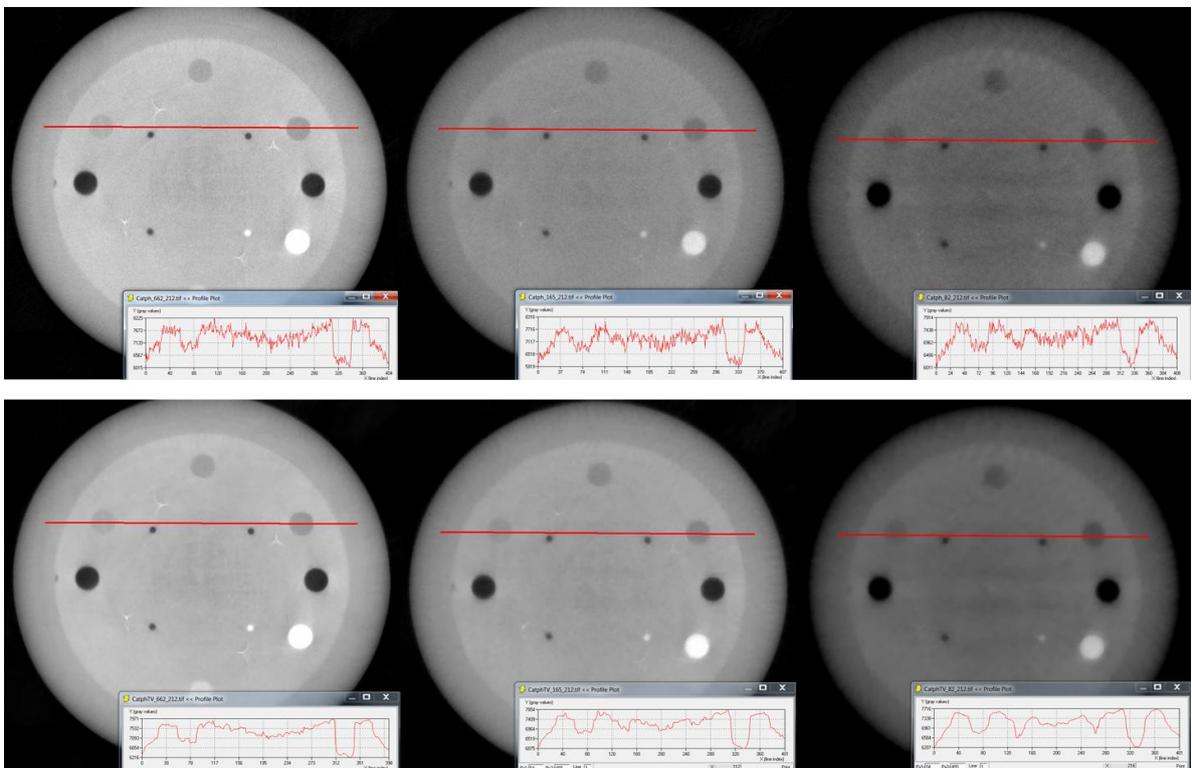
$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} * [1 + \lambda^{(k)} \sum_i A_{ij} / A_{i+} (p_i - (A^i, \mathbf{x}^{(k)})) / (A^i, \mathbf{x}^{(k)})], \quad (5)$$

где  $0 < \lambda^{(k)} < 1$ .

Алгоритм коррекции, представленный формулой (5), является практически таким же удобным для распараллеливания с помощью графической библиотеки *OpenGL*, как и известный алгоритм *SART* [3], описываемый формулой (4). Прямое проецирование трехмерной текстуры, содержащей изображение текущего приближения, осуществляется так же, как и для алгоритма *SART* [4]. Отличие заключается в том, что для его реализации приходится ввести две дополнительные двухмерные текстуры. В одной из них содержится корректирующее изображение, а во второй – изображение корректируемого слоя трехмерной текстуры. Корректирующее изображение сначала перспективно проецируется [5] на место, где находится корректируемый слой, а потом перемножается с изображением корректируемого слоя и записывается на его место.

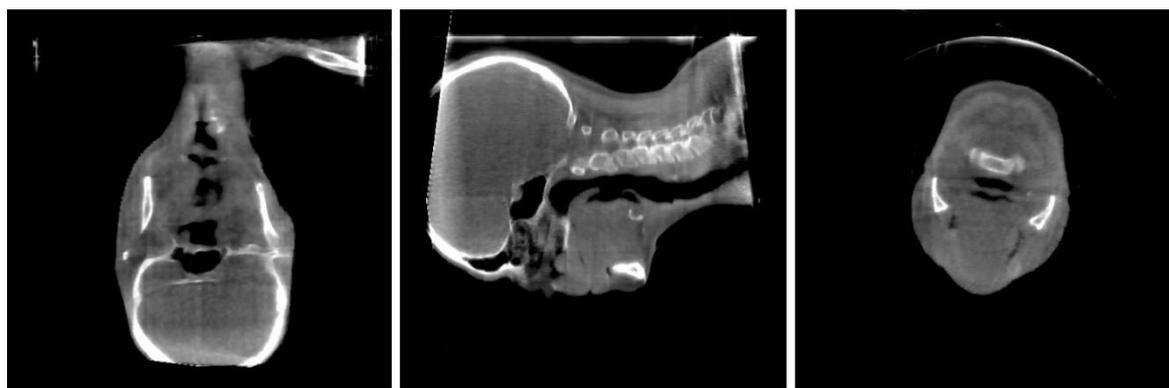


а)



б)

Рис. 1. Результаты реконструкции фантома *CATPHAN*: а) в месте расположения вставки *CTP528*; б) в месте расположения вставки *CTP404*

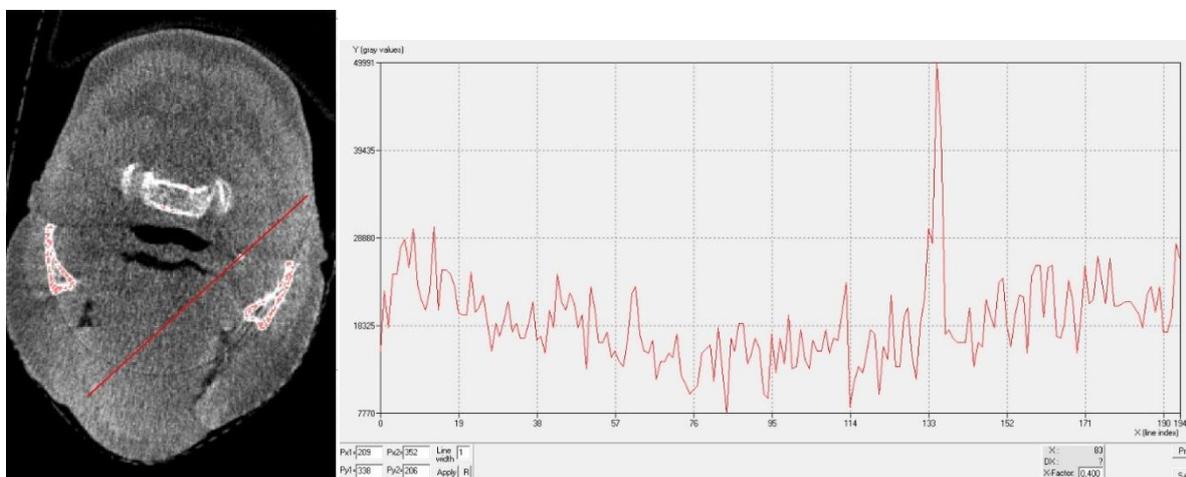


а)

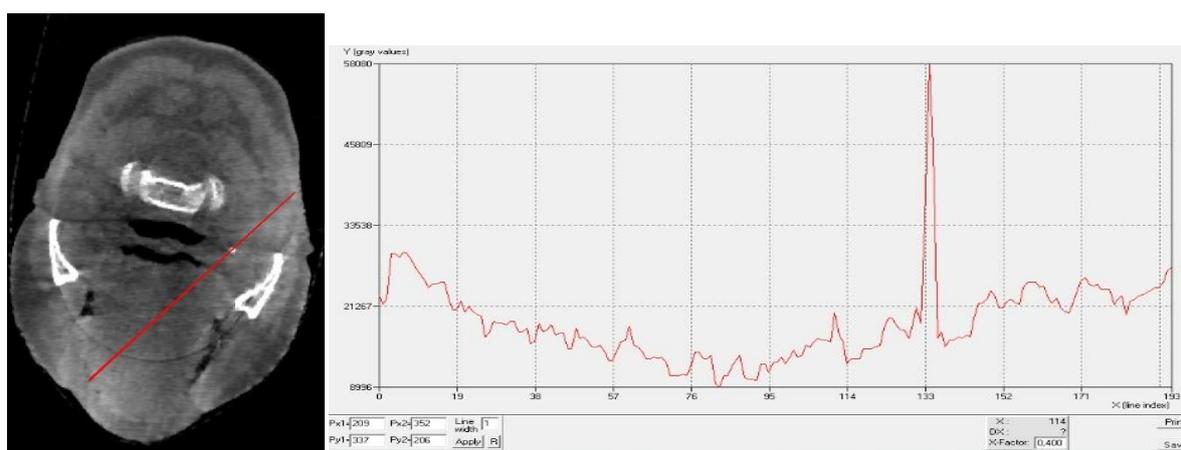
б)

в)

Рис. 2. Ортогональные сечения изображения объекта «голова – шея»: а) XOY; б) YOZ; в) XOZ. Реконструкция по 90 проекциям с TV-регуляризацией



а)



б)

Рис. 3. Изображение 256-го фронтального слоя, реконструированного для пациента: а) без регуляризации; б) с TV-регуляризацией. Слева – изображение вместе с линией, справа – профиль

Для перемножения двумерных текстур было использовано наложение текстуры на объект с режимом *GL\_MODULATE*.

Помимо мегавольтного рентгеновского источника в многофункциональном ускорителе Электа Синерджи имеется киловольтный источник рентгеновского излучения, предназначенный для осуществления конической томографической реконструкции изображений пациентов с целью их правильного позиционирования перед проведением лучевой терапии. Рентгеновские проекции физического фантома *САТФАН* и объекта «голова – шея» для нескольких пациентов были получены с помощью киловольтного рентгеновского источника.

Минимизация *TV* [6] для трехмерного изображения осуществлялась на обычном процессоре в перерыве между корректирующими итерациями, которые были реализованы на графической видеокарте *NVIDIA GeForce GTX 470* с 1280 МБ текстурной памяти.

## 2. Результаты исследования

Все проекции были получены на линейном ускорителе последнего поколения Электа Синерджи. Параметры для рентгеновской съемки реального физического фантома *САТФАН 600*: напряжение на трубке – 120 кВ, анодный ток – 40 мА, время экспозиции – 40 мс. Всего были получены 662 рентгеновские проекции с размерностями 1024×1024 пикселей в угловом диапазоне 360°. Они были дополнительно равномерно разбиты на три группы: 331, 165 и 82 проекции, лежащие в том же угловом диапазоне. Параметры для рентгеновской съемки объекта «голова – шея»: напряжение на трубке – 100 кВ, анодный ток – 10 мА, время экспозиции – 10 мс. Для реконструкций использовались 373 проекции и 90 проекций, расположенных в угловом диапазоне 201°.

Результаты томографических реконструкций показаны на рис. 1–3. На рис. 1, а изображены поперечные сечения реконструированного трехмерного изображения фантома *САТФАН 600* в месте расположения специальной вставки *СТР528* для 82, 165 и 662 проекций соответственно, на рис. 1, б – поперечные сечения реконструированного трехмерного изображения фантома *САТФАН 600* в месте расположения специальной вставки *СТР404* также для 82, 165 и 662 проекций. Верхний ряд соответствует реконструкциям, осуществленным без регуляризации, а нижний ряд – реконструкциям с *TV*-регуляризацией. На рис. 2 изображены характерные поперечные сечения реконструированного трехмерного изображения объекта «голова – шея» для 90 проекций, а на рис. 3 – для 373 проекций. Для демонстрации эффективности использования *TV* в качестве априорного поддерживающего функционала на всех изображениях показаны профили сигнала вдоль выбранных направлений, заданных прямыми линиями. Видно, что шумовая составляющая значительно ослаблена. В то же время границы между различными областями остались четкими. Размерности всех реконструированных трехмерных изображений – 512×512×512 вокселей.

## Заключение

В статье исследован предложенный ранее автором параллельный эвристический алгоритм реконструкции для круговой схемы сканирования в коническом пучке рентгеновских лучей [2]. Проведена серия реконструкций на специальном физическом фантоме *САТФАН*. Полученные результаты свидетельствуют о высоком быстродействии и хорошем контрасте реконструированных изображений данного фантома. Дополнительная проверка алгоритма проведена с помощью реконструкций по рентгеновским проекциям реальных пациентов. Было показано, что он может быть использован в качестве альтернативы алгоритму Фельдкамп, который применяется на линейном ускорителе Электа Синерджи. Установлено, что при регуляризации алгоритма методом *TV* значительно уменьшается уровень шума в реконструируемом трехмерном изображении и улучшается его качество с одновременным сохранением четких границ.

## Список литературы

1. Zolotarev, S.A. 3D Iterative Conical and Spiral Tomography, Laminography, Tomographic Mammography and Tomosynthesis using OpenGL / S.A. Zolotarev, V.L. Vengrinovich, M.A. Mirzavand // NDT Days 2015. Annual Conf. of the Bulgarian Society for NDT, Workshop on Advanced

Techniques in Non Destructive Testing Part II. New devices for NDT and procedures for its application. – Sozopol, Bulgaria, 2015.

2. Lange, K. Globally Convergent Algorithms for Maximum a Posteriori Transmission Tomography / K. Lange, J.A. Fessler // IEEE Trans. on Image Processing. – 1995. – Vol. 8, № 10. – P. 1430–1438.

3. Andersen, A.H. Algebraic reconstruction in CT from limited views / A.H. Andersen // IEEE Trans. Medical Imaging. – 1989. – Vol. 8. – P. 50–55.

4. Венгринович, В.Л. Итерационные методы томографии / В.Л. Венгринович, С.А. Золотарев. – Минск : Беларуская навука, 2009. – 227 с.

5. Fast shadows and lighting effects using texture mapping / M. Segal [et al.] // Proc. of the 19th Annual Conf. on Computer Graphics and Interactive Techniques (SIGGRAPH'92). – N.Y., 1992. – Vol. 26. – P. 249–252.

6. Jia, R.Q. A fast algorithm for the total variation model of image denoising / R.Q. Jia, H.Q. Zhao // Adv. Comput. Math. – 2010. – Vol. 33.– P. 231–241.

Поступила 21.10.2015

*Белорусский национальный  
технический университет,  
Минск, пр. Независимости, 65  
e-mail: mohsen.mirzavand@yahoo.co.uk*

**М.А. Mirzavand**

## **PARALLEL ITERATIVE RECONSTRUCTION OF PHANTOM CATPHAN ON EXPERIMENTAL DATA**

The principles of fast parallel iterative algorithms based on the use of graphics accelerators and OpenGL library are considered in the paper. The proposed approach provides simultaneous minimization of the residuals of the desired solution and total variation of the reconstructed three-dimensional image. The number of necessary input data, i. e. conical X-ray projections, can be reduced several times. It means in a corresponding number of times the possibility to reduce radiation exposure to the patient. At the same time maintain the necessary contrast and spatial resolution of three-dimensional image of the patient. Heuristic iterative algorithm can be used as an alternative to the well-known three-dimensional Feldkamp algorithm.

## БИОИНФОРМАТИКА

УДК 004.048

Н.А. Новоселова, И.Э. Том

## МЕТОД ПОСТРОЕНИЯ КЛАСТЕРОВ ГЕНЕТИЧЕСКИХ ДАННЫХ

*Предлагается метод построения кластеров (функциональных модулей) генетических данных, основанный на использовании рандомизированных матриц. Для построения кластеров выполняется выделение и анализ главных компонент матрицы корреляций генных профилей. В качестве конечных выбираются главные компоненты, которые соответствуют собственным значениям, значимо отличающимся от полученных при анализе случайным образом сгенерированной корреляционной матрицы (рандомизированной). В сравнительном вычислительном эксперименте с аналогами метод показал свое преимущество в возможности выделять статистически значимые кластеры малых и больших размеров, способности отфильтровывать неинформативные признаки, а также получать биологически интерпретируемые функциональные модули, адекватные реальной структуре данных.*

**Введение**

Новые технологии секвенирования генных последовательностей, которые позволяют получить большое количество информации о геноме различных организмов, способствовали развитию различных компьютерных методов, предназначенных для поиска схожих последовательностей ДНК, предсказания структуры белков и функциональных свойств генов. Особенно бурно данное направление стало развиваться после появления таких геномных технологий, как олигонуклеотидные и кДНК-микрочипы (кДНК – комплементарная ДНК). Типичный набор данных экспрессии генов, полученный с использованием геномной технологии, включает, как правило, сотни образцов для тысяч и десятков тысяч генов, которые содержат большой объем скрытой информации. Для ее обнаружения необходимо использовать специальные методы, которые в дополнение к анализу ДНК-последовательностей, заключающемуся в поиске генов и их положения в последовательности, позволяют выделить функциональные модули схожих генных профилей. Гены, относящиеся к одному и тому же кластеру, обычно отвечают за определенный физиологический процесс или относятся к одному и тому же молекулярному комплексу.

**1. Обзор существующих методов**

Для выделения функциональных модулей в данных генной экспрессии были предложены различные методы машинного обучения, которые позволяют строить клеточные сети. Наиболее широко используемыми являются методы построения булевых сетей, байесовских сетей, иерархический кластерный анализ, кластерный метод  $k$ -средних, самоорганизующиеся карты Кохонена (SOM) и ассоциативные корреляционные сети.

Метод построения булевых сетей [1] использует упрощенное представление генных связей, где 1 кодирует наличие связи, а 0 – ее отсутствие. Байесовские сети позволяют графически представить зависимость между генами на основе оценки условных вероятностей. При иерархической кластеризации [2] выполняется итерационный процесс группировки генов, когда первоначально группируются гены с большим значением корреляции. Однако этот алгоритм может сходиться к локальному минимуму, так как объединение генов в кластеры происходит в прямом направлении и решение, принятое на некоторой итерации алгоритма, не может быть изменено на последующих итерациях. SOM [3] представляет собой нейросетевой кластерный алгоритм, для которого требуется задание начального числа кластеров, что почти всегда затруднительно. Ассоциативные корреляционные сети широко используются для идентификации клеточных сетей, поскольку обладают вычислительной эффективностью и способностью

справляться с особенностями данных микрочипов (зашумленностью, высокой размерностью и значительным количеством пропусков). Недостатком таких сетей является произвольным образом задаваемые пороговые значения для установления связей в сети, что приводит к субъективности в определении структуры и топологии сетей. Известны различные программные комплексы кластеризации и визуализации данных микрочипов, например: affy, cclust, cluster, mcluster, hybridHclust, SOM-пакет в среде R [4], Bioconductor [5] и Cluster3.0/Tree view [6], веб-системы, такие как cyberT [7], SNOMAD [8]. Недостатки описанных выше методов не позволяют выявлять наиболее значимые кластеры и отфильтровывать шумы в данных, поэтому задача построения транскрипторных сетей и обнаружения несмещенных оценок биологических связей в данных генной экспрессии продолжает оставаться актуальной и исследованиям в этом направлении уделяется большое внимание.

Последующие разделы посвящены подробному описанию предлагаемого метода, итерационной процедуры процесса выделения функциональных модулей, процедуры оценки стабильности полученных кластеров, а также результатам тестирования метода на наборах искусственно сгенерированных и реальных данных.

## 2. Описание метода построения кластеров генетических данных

Определим значение экспрессии гена  $i$ ,  $i = 1, \dots, N$ , в нескольких экспериментах как

$$W_i(s) = \ln \left( \frac{Es_i(s)}{Ec_i(s)} \right),$$

где  $Es_i(s)$  – значение экспрессии гена  $i$  в эксперименте  $s$ ,  $s = 1, \dots, K$ ;  $Ec_i(s)$  – контрольное значение. Для того чтобы учесть различные уровни экспрессии для разных генов, проводится стандартизация генных профилей:

$$w_i(s) = \frac{W_i(s) - \bar{W}_i}{\sigma_i},$$

где  $\sigma_i = \sqrt{W_i^2 - (\bar{W}_i)^2}$  – стандартное отклонение переменной  $W_i$ , а  $\bar{W}_i$  – среднее значение по различным экспериментам для гена  $i$ . Используя стандартизованную матрицу  $M$  размерности  $N \times K$ , рассчитаем матрицу кросс-корреляции  $C$ :

$$C = \left( \frac{1}{K} \right) MM^T.$$

Коэффициент корреляции Пирсона  $C_{xy}$  между генами  $x$  и  $y$ , заданными  $K$ -мерным вектором, может быть также рассчитан с использованием формулы

$$C_{xy} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{(K-1)s_x s_y},$$

где  $s_x$ ,  $s_y$  – стандартные отклонения.

Значение коэффициента корреляции изменяется в диапазоне от  $-1$  до  $+1$ . При  $C_{ij} = 0$  корреляция между генами отсутствует. Однако оценить точное значение коэффициента корреляции генных профилей затруднительно в связи с особенностями данных микрочипов. Кросс-корреляция между любой парой генов не всегда является константой, а может изменяться во времени или при различных условиях экспериментов. Кроме того, в условиях ограниченного количества экспериментов на значение корреляции оказывают влияние ошибки эксперимента,

или экспериментальный шум. Чтобы отфильтровать случайную компоненту при оценке значений матрицы кросс-корреляции, проводится сравнительный анализ собственных значений данной матрицы и значений матрицы, случайным образом сгенерированной (рандомизированной). Статистические показатели, соответствующие рандомизированной матрице, не принимаются в дальнейшее рассмотрение и помечаются как шум. Собственные значения исходной матрицы, отличающиеся от таковых для рандомизированной матрицы, рассматриваются в качестве отправных для дальнейшего анализа и построения транскрипторной сети.

Для того чтобы определить значимые для дальнейшего анализа собственные значения кросс-корреляционной матрицы, анализируются вероятностные распределения  $P^C(\lambda)$  и  $P^R(\lambda)$  собственных значений  $\lambda_i$  соответственно исходной матрицы кросс-корреляции  $C$  и случайным образом сгенерированной матрицы  $R$ . Собственные значения сортируются по возрастанию, т. е.  $\lambda_i < \lambda_{i+1}$ . На рис. 1 показаны распределения  $P^C(\lambda)$  и  $P^R(\lambda)$  для набора данных по анализу клеточного цикла дрожжевых грибов [9]. Из рисунка следует, что множество собственных значений матрицы  $C$  находится в диапазоне значений  $[\lambda_-, \lambda_+]$ , рассчитанных для матрицы  $R$ , но с несколькими значениями, выходящими за пределы нижней  $\lambda_-$  и верхней  $\lambda_+$  границ. Эти собственные значения и соответствуют реальной информации о корреляции между признаками.

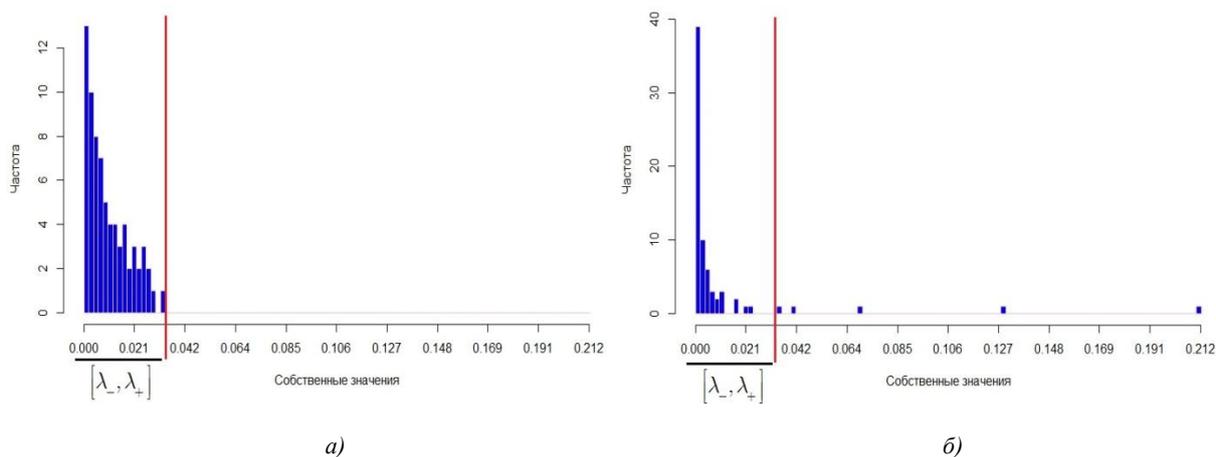
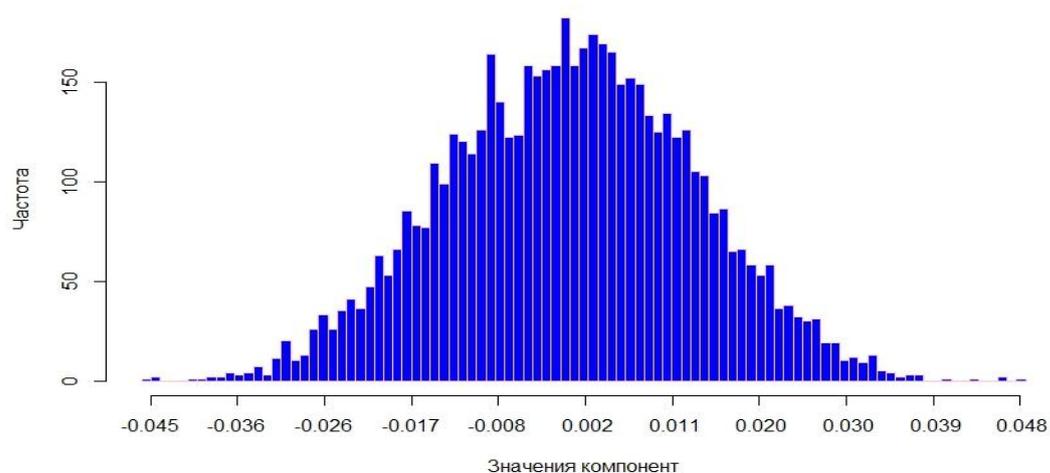


Рис. 1. Гистограммы плотностей распределения собственных значений:  
а) рандомизированная матрица; б) реальная корреляционная матрица

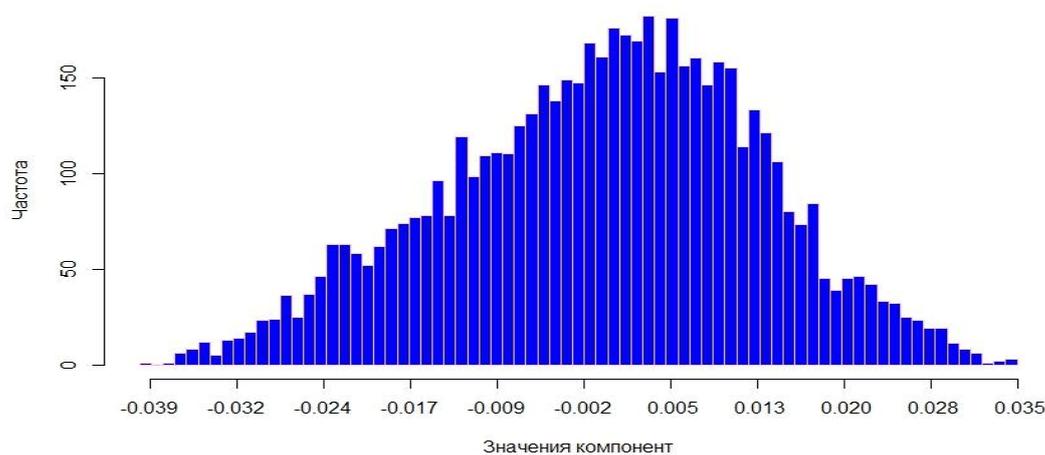
Таким образом, используя собственные значения, можно отделить реальную корреляцию от случайной и выполнить процесс фильтрации шума, что особенно актуально при анализе данных, полученных с помощью технологии микрочипов. Эти данные, как правило, содержат малое число экспериментов по отношению к количеству генов, для которых выполняется измерение экспрессии. Экспериментально подтверждено, что при увеличении количества экспериментов или временного ряда собственные значения корреляционной матрицы все больше отклоняются от диапазона соответствующих значений, полученных для рандомизированной матрицы. Предполагается, что при увеличении числа экспериментов  $K$  удастся выявить наиболее трудноразличимые группы генов, которые невозможно определить с помощью стандартных кластерных методов. Однако на практике выполнение большого количества экспериментов является затратным по времени и требует привлечения значительных материальных ресурсов. Для построения функциональных модулей генов можно использовать только собственные значения, которые выходят за диапазон собственных значений рандомизированной матрицы. Эти функциональные модули и представляют собой кластеры генов, которые выделены из зашумленных данных. Величина дисперсии, приносимая каждым собственным вектором, определяется как отношение соответствующего собственного значения к сумме всех собственных значений корреляционной матрицы. Как и при анализе глав-

ных компонент, где главные факторы объясняют большую часть вариации в данных, в предложенном методе только большие собственные значения и соответствующие собственные векторы рассматриваются для анализа групп генов. Остальные собственные значения являются неинформативными.

Рассмотрим распределение компонент собственных векторов, соответствующих информативным и неинформативным собственным значениям. Пусть имеется  $N$  собственных векторов  $u^i$ ,  $i=1, \dots, N$ . Каждый собственный вектор  $u_i$  имеет  $N$  компонент, соответствующих генам анализируемой матрицы. Вероятностное распределение компонент собственного вектора для собственных значений реальной корреляционной матрицы и рандомизированной матрицы, заданной гауссовым распределением с нулевым средним значением и единичной дисперсией, показано на рис. 2.



а)



б)

Рис. 2. Гистограммы распределения компонент собственного вектора, соответствующего:  
а) собственному значению в диапазоне значений рандомизированной матрицы;  
б) выходящему из диапазона  $[\lambda_-, \lambda_+]$  собственному значению реальной матрицы

Вероятностное распределение компонент собственного вектора  $\lambda_k$  из диапазона  $\lambda_- \leq \lambda_k \leq \lambda_+$  собственных значений рандомизированной матрицы для данных по анализу клеточного цикла дрожжевых грибов хорошо согласуется с нормальным распределением (рис. 2, а).

Распределение компонент собственного вектора, соответствующего собственному значению, выходящему из диапазона, отличается от распределения Гаусса (рис. 2, б). Вычислительные эксперименты показали, что при приближении собственного значения к диапазону значений рандомизированной матрицы их распределение стремится к нормальному.

После того как определены собственные векторы, которые соответствуют информативным собственным значениям (выходящим из диапазона значений для рандомизированной матрицы), компоненты векторов преобразуются в весовые коэффициенты путем перемножения их значений на квадратный корень из соответствующих собственных значений. Каждый собственный вектор далее представляет собой один кластер генов. Большее значение весового коэффициента означает большее влияние соответствующего гена на формирование собственного вектора (другими словами, доминирующую позицию соответствующего гена в кластере). Чтобы упростить структуру собственного вектора и облегчить интерпретацию кластера генов, можно применить процедуру ортогонального вращения собственных векторов. Для выполнения вращения используется метод VARIMAX [10], который позволяет преобразовать основные координатные оси таким образом, чтобы каждый собственный вектор содержал меньшее количество больших значений весовых коэффициентов и большее количество нулей или малых значений весовых коэффициентов. С точки зрения биологии это означает, что каждый ген является составной частью кластеров или оказывает влияние только на малое их количество и каждый кластер состоит из меньшего количества доминирующих генов, чем до вращения. Матрица вращения  $R$  в общем виде определяется как

$$R = \begin{bmatrix} \cos \theta_{i,i} & \cos \theta_{i,j} \\ \cos \theta_{j,i} & \cos \theta_{j,j} \end{bmatrix},$$

где  $\theta_{i,j}$  – угол вращения, преобразующий старые оси в новые.

В предлагаемом методе реализована итерационная процедура выделения функциональных модулей генетических данных. Согласно процедуре функциональные модули выделяются сначала из всего анализируемого набора данных, а затем из каждого выделенного генного кластера. Результатом такой процедуры является иерархическая кластерная структура, подобная структуре, получаемой стандартными методами иерархической кластеризации. Отличием является то, что на каждом последующем шаге процедуры в рассмотрение принимается не все содержимое кластера, а только его часть, соответствующая наиболее значимым собственным значениям корреляционной матрицы.

### 3. Оценка стабильности кластеров

Оценка стабильности выделенных кластеров основана на оценке собственных значений корреляционной матрицы. Новые значения координат экспериментов  $s$  рассчитываются с использованием компонент собственных векторов  $u^i$ , где каждая из координат  $z^i(s)$  представляет собой значение эксперимента в пространстве собственных векторов:

$$z^i(s) = \sum_{k=1}^N u_k^i W_k(s).$$

Стабильность  $i$ -го кластера генов может быть оценена через дисперсию всех экспериментов  $z^i$  по координате собственного вектора  $i$ . Дисперсия напрямую ассоциируется с соответствующим собственным значением:

$$Stab(u^i) = \sigma^2(z^i) = (u_i)^T C u^i = \lambda_i,$$

где  $i=1, \dots, N$ .

Таким образом, генный кластер, порожденный собственным вектором с бóльшим собственным значением, является менее стабильным, значение дисперсии соответствует согласованности генов в кластере по всем экспериментам и предоставляет дополнительную информацию о качестве кластера и его свойствах.

#### 4. Результаты тестирования метода на искусственно сгенерированном наборе данных

Для тестирования предлагаемого метода построения кластеров генетических данных, основанного на теории рандомизированных матриц, и проведения его сравнительного анализа с известными методами разработана методика генерации случайных наборов данных, которые имитируют данные экспрессии генов. В соответствии с методикой данные генерируются по следующим правилам: 1) согласно нормальному распределению генерируются несколько многомерных кластеров генов, причем предполагается, что отдельные признаки (случаи или элементы временного ряда) являются независимыми; 2) генерируемые кластеры имеют различные профили (за счет искусственного разбиения признаков на несколько интервалов, значения внутри каждого из которых сгенерированы согласно индивидуальным вероятностным распределениям); 3) размеры кластеров варьируются от 10 до 100 элементов; 4) генерируются несколько кластеров шумов для проверки способности алгоритма выделять и исключать из рассмотрения незначимые генные кластеры.

Использованный для тестирования сгенерированный набор данных состоял из четырех профилей генных кластеров (табл. 1). Каждый кластер содержал различное количество генов, всего 85 информативных генов. Случайным образом согласно нормальному распределению генерировались дополнительно 50 генов, которые были распределены по 100 объектам согласно  $N(0,1)$  и представляли собой шум в данных. Все признаки были разделены на пять классов (по 20 объектов каждого класса).

Таблица 1

Параметры сгенерированного набора данных

Кластер	Средние значения профилей генов для пяти классов	Стандартное отклонение	Количество генов
1	1, -1, -1, -1, -1	0,2	5
2	-1, 1, -1, -1, -1	0,2	10
3	-1, -1, 1, -1, -1	0,2	20
4	-1, -1, -1, 1, -1	0,2	50

В результате применения метода были выявлены все четыре функциональных модуля генов и отсеян шум. На рис. 3 представлены четыре профиля полученных кластеров.

Таким образом, экспериментально удалось показать, что предложенный метод способен выявлять кластеры, имеющие разные размеры, и отфильтровывать неинформативные признаки. Rand-индекс [11] качества кластеризации равен единице.

Для выполнения сравнительного анализа результатов, полученных предложенным методом и методом  $k$ -средних, был использован R-пакет `sclust`, который позволяет итерационно смещать центры кластеров к средним значениям областей соответствующей диаграммы Вороного. В качестве расстояния между объектами используется евклидово расстояние. Основным недостатком метода  $k$ -средних является необходимость изначального определения числа кластеров, которое в большинстве случаев заранее неизвестно. Для проведения эксперимента количество кластеров было определено равным пяти в предположении, что все неинформативные гены определяют отдельный кластер.

В результате метод  $k$ -средних оказался неспособным выявить все информативные кластеры. Было получено пять кластеров, из которых кластер 1 являлся объединением исходных кластеров 1 и 2 (см. табл. 1), кластер 2 соответствовал исходному кластеру 3, кластер 3 соответствовал исходному кластеру 4 и неинформативные гены были разбиты на два оставшихся кластера (табл. 2). Rand-индекс качества кластеризации равен 0,79.

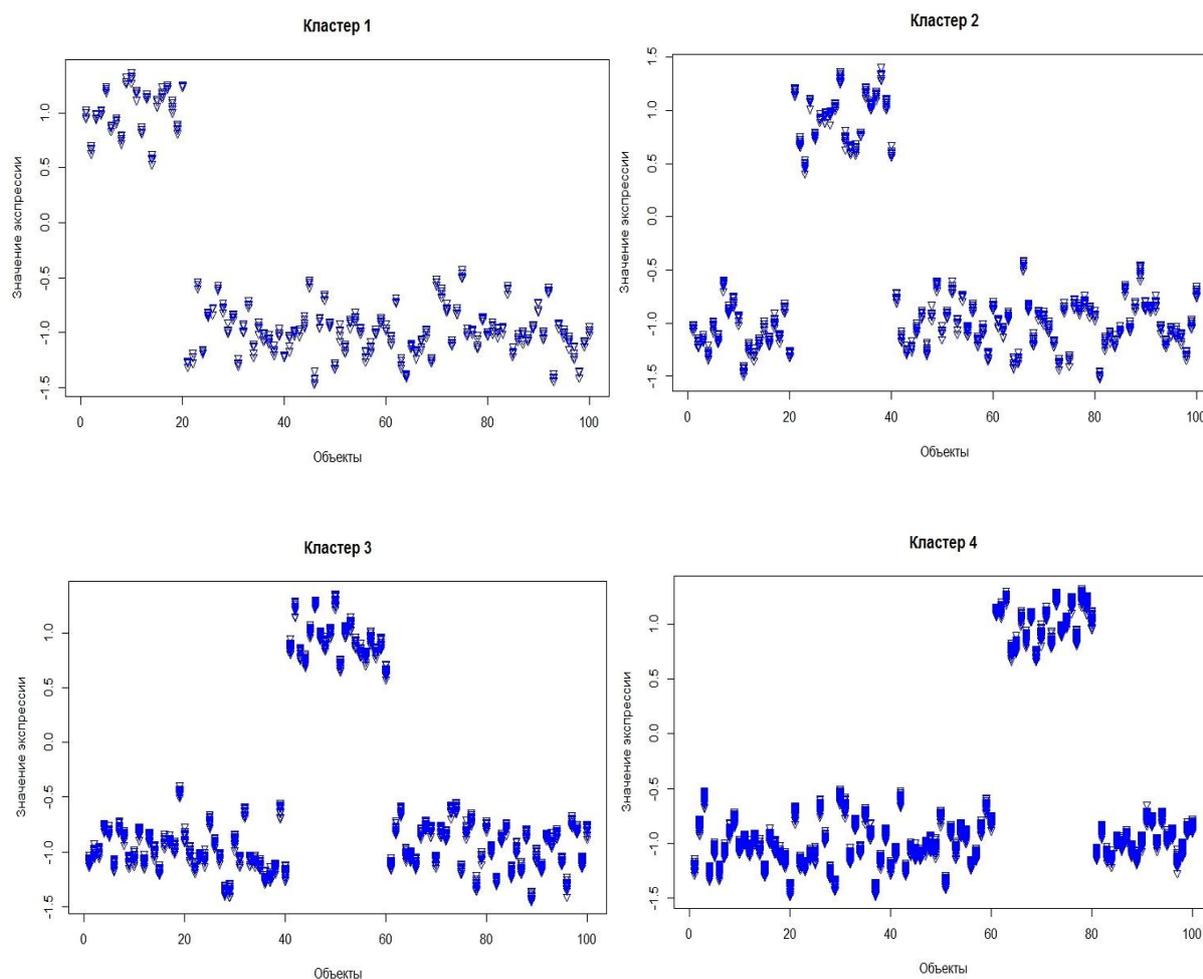


Рис. 3. Профили кластеров, полученных с использованием предложенного метода

Таблица 2

Результаты кластеризации методом  $k$ -средних

Кластер	Значения генов исходного набора данных	Кластеры генов исходного набора данных	Количество генов
1	1–15	1, 2	15
2	16–35	3	20
3	36–85	4	50
4	86–96, 100, 103–105, 107, 109, 111–113, 117, 118, 122, 123, 125, 126, 128, 130	Неинформативные гены	28
5	97–99, 101, 102, 106, 108, 110, 114–116, 119–121, 124, 127, 129, 131–135	Неинформативные гены	22

Второй сравнительный эксперимент на том же наборе данных был проведен с использованием гибридного кластерного метода *hybridHclust* [12], представляющего собой иерархический метод кластеризации, широко применяемый для анализа данных генной экспрессии и позволяющий выявить иерархическую организацию кластерной структуры. Метод позволяет компенсировать ряд недостатков базовых методов иерархической кластеризации, объединяя кластеризацию снизу вверх с кластеризацией по нисходящему принципу. Для проведения эксперимента использовался R-пакет *hybridHclust*, а для вычисления близости объектов приме-

нялось евклидово расстояние. В результате было определено 14 кластеров, среди которых имеются все четыре исходных кластера из табл. 1. В кластер 4 кроме исходного кластера вошел и неинформативный ген 113. Кроме того, были получены 10 дополнительных кластеров, которые являются неинформативными в связи с тем, что соответствующие им гены имеют случайным образом сгенерированный профиль (табл. 3).

Таблица 3

Результаты кластеризации методом hybridHclust

Кластер	Значения генов исходного набора данных	Кластер	Значения генов исходного набора данных
1	1–5	8	90, 96
2	6–15	9	92, 134
3	16–353	10	104, 111
4	36–85, 113	11	107, 130
5	94, 114	12	89, 109
6	100, 101	13	102, 110
7	105, 133	14	112, 131

Можно сделать вывод, что метод hybridHclust в отличие от предложенного не позволяет точно определить исходные кластеры и отделить шум от информативных признаков.

##### 5. Результаты тестирования метода на наборе данных по анализу клеточного цикла дрожжевых грибов

Набор данных дрожжевых грибов [9] является результатом применения технологии микрочипов для анализа клеточного цикла. Набор состоит из значений экспрессии 2500 генов, полученных в результате 79 экспериментов. Дрожжи представляют собой классический модельный организм как для исследований фундаментальных генетических проблем, так и для изучения наследственных болезней человека в связи с тем, что многие из основных клеточных процессов консервативны и одинаковы у дрожжей и у человека. Например, многие из известных сейчас молекул и метаболических путей, вовлеченных в формирование злокачественных новообразований (например, клеточный цикл и его контролирование, репарация ДНК и т. д.), впервые были открыты и исследованы на дрожжах. Набор данных содержит результаты экспериментов с различными штаммами грибов для выявления генов, ассоциированных с различными этапами клеточного цикла.

С помощью предложенного метода из набора данных были выделены 20 функциональных модулей, каждый из которых характеризуется множеством генов со схожими профилями экспрессии. В состав каждого модуля вошли гены, имеющие наибольшие весовые коэффициенты, соответствующие компонентам значимых собственных векторов. Изначально предполагалось, что гены, относящиеся к одному модулю, вовлечены в похожие биологические процессы. Для определения биологических процессов для генных кластеров была использована веб-среда YeastMine [13], которая позволяет получить генетическую информацию для дрожжевых грибов *Saccharomyces cerevisiae*. Ее применение обеспечивает возможность поиска по набору генов степени их вовлечения в биологический процесс, сигнальный путь, GO-аннотацию генов и т. д. В табл. 4 представлен ряд выделенных функциональных модулей, которые соответствуют процессам биогенезиса протеинов, репликации и восстановления ДНК, энергетического метаболизма, деградации протеинов, сворачивания протеинов, а также протеинам теплового шока, циклу Кребса, механизму разложения аллантиина и регуляции гистонов.

Согласно разработанной процедуре итерационной кластеризации модули, состоящие из большого числа генов, были повторно кластеризованы для выявления подгрупп в составе модуля. Для наибольшей группы, содержащей 230 генов, были выделены два подмодуля, соответствующие процессам гликолиза и клеточного цикла. Таким образом удалось показать, что результаты применения метода позволяют выявить функциональные модули генов, являющиеся информативными и с биологической точки зрения.

Таблица 4

Пример выявленных функциональных модулей генов, аннотированных согласно YeastMine

Функциональный модуль	Гены	Функциональный модуль	Гены
Биогенезис протеинов	RPL8B, RPL34B, RPS19A, RPS26B, RPL7B, RPL14A, RPS6B, RPL33B, RPS1B, RPL9A, RPL11B, RPS24A, RPL26B, RPS15, RPL24A, RPL12B, RPS0A, RPS19B, RPL19B, RPL27B, RPS29B, RPL13A, RPS4B, RPL33A, RPS29A, RPL18B, RPL11A,...	Протеины теплового шока	SSA2, SSA1, ECM10, SSA4, KAR2, STI1, SIS1, SUR2, YME1, SRP21, FCY1, SUI1, VAM3, PFD4
		Регуляция гистонов	HHT1, HTA1, HTB1, HNF2, HTA2, HHT2, HNF1, HTB2, HHO1
Репликация и восстановление ДНК	CDC45, KIM2, MSH2, DUN1, POL30, SMC3, RFA2, RNR1, SWE1, POL2, POL32, RFA1, SEN34, PIF1, SPH1, OGG1, CDC21, RHC18, RFC3, RAD27, TUB4, ZDS2, ASF1, HCM1, MSH6, RFC5, SAS2, BNI4, CDC2, PMS1, ARP1, POL1, TOP1, RAD51, FCP1, PMT5, PCH1, RNR3, MRE11, ALG2, PRI1, RFA3, BUD2, DUT1, ASF2, ADK2, CDC9, ECM25, RFC4	Цикл Кребса	CYT1, SDH2, MDH1, SDH4, QCR6, ACH1, COR1, SDH3, RIP1
		Механизм разложения аллантаина	DAL2, DAL1, DAL3, DAL5, MEP2
Деградация протеинов	SCL1, PUP2, PRE2, PRE5, PRE1, PRE3, RPN6, PRE4, RPT6, PRE9, RPT1, RPN9, RPT4, RPN10, RPN11, PRE7, PUP1, RPN3, PRE8, PRE10, UFD1, UMP1, PRE6, RPN12, PUP3, RPN7, RPT3, RPN2, RPT5, GSH2, QRI8, STE24	Энергетический метаболизм	MRPL9, MRPL13, ATP12, PET123, COX12, MRPL35, MRP2, ATP11, MRP17, DBI56, QCR8, MRP51, PPA2, HXT4, COX17, RML2, MRPS5, MRPL38, MRP7, MRPL25, ATP14, MSS51, FMC1, MSM1, MRPL8, CYT2, COX14, MEF1, MRPL32, MSN4, CBP4, PHB1, MSF1, MRPL16, MRPL6, IMG2, MRP13, NAM2, MAS1, MRPL36, NAM9, CYC3, KIM4, CBP3, STS1, COX10, MBA1, MRPL31, ECM19, STP4, CAF16, COX13, CAP1

Для сравнения набор данных по анализу клеточного цикла дрожжевых грибов был кластеризован с использованием метода *k*-средних. Начальное количество кластеров было задано равным 20, что соответствовало результату, полученному с использованием предложенного метода. Метод *k*-средних позволил определить группу генов, связанных с процессом биогенеза протеинов, однако даже эта группа была определена не полностью и часть связанных с данным процессом генов была перераспределена в другие кластеры. Кроме того, метод *k*-средних был не способен выявить малые функциональные модули, такие как гистоны и цикл Кребса, так как целевая функция данного метода предполагает построение кластеров сравнимых размеров и малые кластеры при этом всегда являются составной частью кластеров большего размера.

Набор данных по анализу клеточного цикла дрожжевых грибов также был кластеризован с помощью метода *hybridHclust* [12]. Полученная в результате вычислительного эксперимента дендрограмма оказалась трудна в интерпретации из-за большого числа анализируемых генов. Размерность большинства полученных кластеров была очень мала – в диапазоне от двух до восьми элементов. Надо отметить, что ряд кластеров, выделенных с помощью метода *hybridHclust*, были биологически интерпретируемы. Например, пять генов, отвечающих за гликолиз, выделены в один кластер; восемь из девяти генов, связанных с регуляцией гистонов, также образовали кластер. Однако при этом оказалось невозможно выделить кластеры больших размеров. Кроме того, полученные кластеры были чувствительны к малым вариациям в данных, что приводило к изменению их состава.

### Заключение

Согласно разработанному методу построения кластеров генетических данных с использованием теории рандомизированных матриц первоначально рассчитываются главные компоненты исходной корреляционной матрицы генных профилей. Для учета дисбаланса между количеством генов и экспериментов, полученных по технологии микрочипов, в качестве конечных выбираются главные компоненты, которые соответствуют ограниченному количеству собственных значений. Величины этих собственных значений значимо отличаются от значений, полученных при анализе рандомизированной корреляционной матрицы  $R$ . Статистические характеристики случайной корреляционной матрицы позволяют отфильтровать шум из исходной анализируемой матрицы данных и выделить статистически значимые функциональные модули.

Сравнительный анализ результатов применения предложенного метода и двух известных методов ( $k$ -средних и иерархической кластеризации *hybridHclust*) показал преимущество предложенного метода выделения информативных кластеров. Кластеризация методом  $k$ -средних не способна была распознать кластеры малых размеров, и, наоборот, результатом иерархической кластеризации стало большое количество кластеров малого размера, некоторые из которых являлись составной частью реальных больших кластеров. Кроме этого, метод *hybridHclust* оказался неспособным отфильтровать неинформативные гены.

Таким образом, сравнительный анализ разработанного метода построения кластеров генетических данных с использованием рандомизированных матриц с наиболее известными методами-аналогами показал, что он позволяет более корректно выявлять структуру в данных генной экспрессии как для больших, так и для малых кластеров, успешно фильтруя при этом неинформативные гены.

### Список литературы

1. Liang, S. REVEAL, a general reverse engineering algorithm for inference of genetic network architectures / S. Liang, S. Fuhrman, R. Somogyi // Pacific Symp. on Biocomputing (PSB'98). – Hawaii, 1998. – Vol. 3. – P. 18–29.
2. Cluster analysis and display of genome-wide expression patterns / M.B. Eisen [et al.] // Proceedings of the National Academy of Sciences of the United States of America. – 1998. – Vol. 95. – P. 14863–14868.
3. Analysis of gene expression data using self-organizing maps / P. Toronen [et al.] // FEBS Letters. – 1999. – Vol. 451. – P. 142–146.
4. The R Project for Statistical Computing. R Foundation for Statistical Computing [Electronic resource]. – 2009. – Mode of access : <http://www.R-project.org>. – Date of access : 10.09.2015.
5. Bioconductor case studies / F. Hahne [et al.]. – Springer Science & Business Media, 2010. – 296 p.
6. Cluster – Cluster analysis and visualization software [Electronic resource]. – 2015. – Mode of access : <http://rana.lbl.gov/EisenSoftware.htm>. – Date of access : 19.08.2015.

7. Cyber-T – microarray analysis web interface from UCI’s Institute for Genomics and Bioinformatics [Electronic resource]. – 2015. – Mode of access : <http://cybert.microarray.ics.uci.edu>. – Date of access : 16.09.2015.
8. SNOMAD – Standardization and normalization of microarray data [Electronic resource]. – 2015. – Mode of access : <http://pevsnerlab.kennedykrieger.org/snomadinput.html>. – Date of access : 12.09.2015.
9. Yeast cell cycle analysis project [Electronic resource]. – 2015. – Mode of access : <http://genome-www.stanford.edu/cellcycle>. – Date of access : 10.04.2015.
10. Varimax – rotation methods for factor analysis [Electronic resource]. – 2015. – Mode of access : <https://stat.ethz.ch/R-manual/R-devel/library/stats/html/varimax.html>. – Date of access : 17.09.2015.
11. Morey, L.C. The measurement of classification agreement: an adjustment to the rand statistic for chance agreement / L.C. Morey, A. Agresti // Educational and Psychological Measurement. – 1984. – Vol. 44. – P. 33–37.
12. Chipman, H. Hybrid hierarchical clustering with applications to microarray data / H. Chipman, R. Tibshirani // Biostatistics. – 2006. – Vol. 7, № 2. – P. 286–301.
13. YeastMine: saccharomyces genome database [Electronic resource]. – 2015. – Mode of access : <http://yeastmine.yeastgenome.org/yeastmine/begin.do>. – Date of access : 06.09.2015.

Поступила 09.10.2015

*Объединенный институт проблем  
информатики НАН Беларуси,  
Минск, Сурганова, 6  
e-mail: novosel@newman.bas-net.by*

**N.A. Novoselova, I.E. Tom**

## **METHOD OF CONSTRUCTION OF GENETIC DATA CLUSTERS**

The paper presents a method of construction of genetic data clusters (functional modules) using the randomized matrices. To build the functional modules the selection and analysis of the eigenvalues of the gene profiles correlation matrix is performed. The principal components, corresponding to the eigenvalues, which are significantly different from those obtained for the randomly generated correlation matrix, are used for the analysis. Each selected principal component forms gene cluster. In a comparative experiment with the analogs the proposed method shows the advantage in allocating statistically significant different-sized clusters, the ability to filter non-informative genes and to extract the biologically interpretable functional modules matching the real data structure.

УДК 004.021

Р.С. Сергеев<sup>1</sup>, И.С. Ковалев<sup>2</sup>, А.В. Тузиков<sup>1</sup>, А. Розенталь<sup>3</sup>, А. Габриэлян<sup>3</sup>**АЛГОРИТМЫ ПОИСКА МУТАЦИЙ ЛЕКАРСТВЕННОЙ УСТОЙЧИВОСТИ  
В ГЕНОМАХ МИКОБАКТЕРИЙ ТУБЕРКУЛЕЗА**

*Предлагается методология полногеномного поиска ассоциаций на примере геномов возбудителей туберкулеза, а также исследуются различные алгоритмы и подходы к оценке вклада мутаций в возникновение и развитие лекарственной устойчивости. Приводятся результаты экспериментов по поиску мутаций лекарственной устойчивости к основным противотуберкулезным препаратам на основании данных о пациентах из Беларуси.*

**Введение**

Стремительное развитие высокопроизводительных методов секвенирования для определения нуклеотидных последовательностей ДНК живых организмов придает значительный импульс биологическим исследованиям и становлению персонализированной медицины. Однако сам по себе прочитанный генетический код не имеет большой практической ценности до извлечения из него нужной информации. Анализ расшифрованных геномных последовательностей зачастую приводит к задачам большой размерности, где число неизвестных параметров измеряется десятками тысяч при относительно небольшом числе доступных наблюдений. Одной из таких задач является поиск мутаций в геномах микроорганизмов, которые непосредственно обуславливают развитие лекарственной устойчивости к применяемым в ходе терапии лекарственными препаратами.

Актуальность задачи состоит в том, что в последнее время все чаще выявляются штаммы *Mycobacterium tuberculosis*, обладающие множественной и широкой лекарственной устойчивостью. Причина возникновения подобных штаммов во многом обусловлена такими факторами, как выбор неверной схемы лечения, нарушение режима химиотерапии, поздняя постановка диагноза [1]. Эти ошибки приводят к последовательному накоплению изменений в генах, прямо или косвенно вовлеченных во взаимодействие с лекарственными субстанциями.

Предлагаемая методология позволяет по набору полных геномов *M. tuberculosis* проанализировать влияние мутаций на развитие резистентности к противомикробным препаратам и на основании этой информации выполнить отбор участков генома для формирования базы данных структурных полиморфизмов. Исследование генетических или геномных маркеров лекарственной устойчивости имеет большое значение для своевременной диагностики форм туберкулеза, а также понимания биологических механизмов становления лекарственной устойчивости и выбора потенциальных мишеней для создания новых противомикробных препаратов.

**1. Постановка задачи**

В настоящем исследовании решается задача полногеномного поиска ассоциаций (genome-wide association study, GWAS), где анализируются мутации – однонуклеотидные полиморфизмы (single nucleotide polymorphisms, SNPs) в последовательностях ДНК микобактерий туберкулеза. Цель этой задачи состоит в идентификации участков генома, мутации в которых максимально влияют на фенотип (наличие либо отсутствие лекарственной устойчивости к определенному препарату).

Пусть на входе имеется набор геномных последовательностей  $S_1, S_2, \dots, S_n$ , таких, что  $S_i = s_{i1}s_{i2}\dots s_{il}$ ,  $s_{ij} \in \{A, T, G, C, -\}$ . Здесь  $A, T, G, C$  обозначают четыре нуклеотида, а «-» означает, что элемент последовательности не определен. Будем считать, что все последовательности выравнены относительно друг друга и некоторой референсной последовательности  $S_0$ . Тогда лю-

бое отличие символа  $s_{ij}$  в  $i$ -й последовательности от символа  $s_{oj}$  в этой же позиции референсной последовательности будем называть однонуклеотидным полиморфизмом (точечной мутацией). Для более компактного представления введем матрицу генотипов  $X$  размера  $n \times m$ , строки которой будут соответствовать последовательностям, а в столбцы попадут только позиции исходного выравнивания, содержащие хотя бы одну мутацию. При этом  $x_{ik} = 1$ , если в  $i$ -й последовательности присутствует мутация в позиции, соответствующей  $k$ -му столбцу матрицы  $X$ , и  $x_{ik} = 0$ , если мутация отсутствует.

Существует несколько линий лекарственных препаратов, использующихся для терапии туберкулеза. Препараты первой линии назначаются в комбинациях для лечения первично инфицированных пациентов, у которых не обнаружено резистентных штаммов *M. tuberculosis*. Препараты второй линии применяются в случае, если установлена лекарственная устойчивость к препаратам первой линии. Информацию о результатах тестов на чувствительность к некоторому лекарственному препарату будем кодировать в виде вектора  $Y$  длины  $n$ , элемент которого  $y_i = 1$ , если установлена лекарственная устойчивость  $i$ -го организма к этому препарату, и  $y_i = 0$  в противном случае.

В работе сравниваются ДНК-последовательности, выделенные из двух групп организмов: лекарственно-устойчивых и лекарственно-чувствительных. В качестве референсной последовательности  $S_0$  был выбран геном штамма H37Rv (который сохраняет полную вирулентность и восприимчив к противотуберкулезным препаратам). Соответственно клинические сведения о пациентах служат источником информации, чтобы разделить последовательности на группы устойчивых и чувствительных к выбранному препарату. Таким образом, проблема поиска значимых мутаций в геномах микроорганизмов может быть определена как задача бинарной классификации по известным бинарным признакам.

## 2. Методы и алгоритмы

Процедура анализа данных включает несколько последовательных шагов (рис. 1). На начальном этапе анализируются новые геномы микроорганизмов, выполняются первичная предобработка и очистка данных. Следующие шаги ориентированы на анализ популяционной структуры, поиск коррелирующих сайтов, а также генов, находящихся под воздействием положительного отбора.



Рис. 1. Процедура анализа геномных данных

На завершающих шагах оценивается вклад мутаций в развитие лекарственной устойчивости, происходит сравнение и аннотирование результатов.

### 2.1. Фильтрация и очистка данных

Для уменьшения размерности задачи на этапе предобработки, а также для улучшения качества искомых решений применяется ряд фильтров, позволяющих сократить число столбцов в матрице генотипов. Наиболее широко используемым фильтром является фильтр частоты редкой аллели (minor allele frequency, MAF). Как правило, если мутация (редкая аллель) встречается менее чем в 1 % организмов, то она считается недостоверной и может быть исключена из дальнейшего анализа. В исследуемых данных лишь порядка 30 % мутаций прошли фильтр MAF.

Чтобы дополнительно уменьшить количество параметров в моделях, была выполнена процедура удаления дубликатов из матрицы генотипов  $X$ . Для этого столбцы матрицы  $X$  были разбиты на группы, состоящие из абсолютно идентичных друг другу столбцов; в конечной же матрице был оставлен один представитель из каждой группы. Если в последующих тестах позиция, соответствующая какому-либо столбцу из выбранной группы, признавалась значимой, то в результирующий список включались все позиции, соответствующие остальным столбцам этой группы.

По итогам первичной предобработки данных была получена матрица генотипов, каждый столбец которой уникален и содержит не менее 1 % различий. После объединения исходных VCF-файлов, их фильтрации по качеству, критерию MAF и удаления дубликатов удалось сократить число тестируемых генетических маркеров с 50 000 в сырых данных до порядка 1000 в обработанной информации.

Отфильтрованные и очищенные данные были сгруппированы в наборы, на основе которых происходило последующее обучение. Каждый набор данных представляет собой матрицу, составленную из части строк матрицы генотипов  $X$  и соответствующих этим строкам элементов вектора фенотипов  $Y$ . Выбор строк осуществлялся в зависимости от тестируемого препарата и используемых условий (например, исследовались генетические маркеры лекарственной устойчивости к офлоксацину при допущении, что все образцы текущего набора обладали чувствительностью к инъекционным препаратам второго ряда).

## 2.2. Исследование популяционной структуры

Основной целью анализа популяционной структуры является установка принадлежности исследуемого организма к уже известным таксономическим группам и проведение кластеризации данных. Как правило, для определения групповой структуры генетических последовательностей используются методы филогенетического анализа, сполиготипирования и метод главных компонент.

В данной работе для уменьшения размерности, группирования данных и расчета поправок для последующих статистических тестов ассоциаций использовался метод главных компонент [2]. Результаты сполиготипирования образцов из полученных групп применялись для определения принадлежности каждой группы к одному из известных семейств *M. tuberculosis*. В качестве входящих данных использовалась матрица генотипов  $X$  с элементами  $x_{ij} \in \{0, 1\}$ , строки которой  $i = \overline{1, n}$  соответствуют организмам, а столбцы  $j = \overline{1, m}$  — упорядоченным позициям (сайтам) выравнивания, содержащим мутации. Элементы матрицы  $X$  предварительно нормировались:

$$\bar{x}_{ij} = (x_{ij} - \hat{x}_j) / \sqrt{\hat{x}_j(1 - \hat{x}_j)},$$

где  $\hat{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij}$  — выборочное среднее для  $j$ -го столбца.

По нормированной матрице  $\bar{X}$  вычислялась ковариационная матрица  $C = \frac{1}{n-1} \bar{X} \bar{X}^T$ .

Согласно сингулярному разложению нормированная матрица генотипов представима в виде  $\bar{X} = USV^T$ . Тогда с точностью до константы выполнено  $C = US^2U^T$ , где  $U$  представляет собой матрицу  $n \times n$ , составленную из собственных векторов матрицы  $\bar{X} \bar{X}^T$ , которые образуют ортонормированный базис. Каждый элемент  $u_{li}$  матрицы  $U$  может быть интерпретирован как «координата»  $i$ -го организма на  $l$ -й оси вариаций (главной компоненте). Для визуализации использовались проекции на первые две главные компоненты (рис. 2), которые отражают популяционную структуру, где генетически наиболее близкие друг другу организмы объединены в группы.

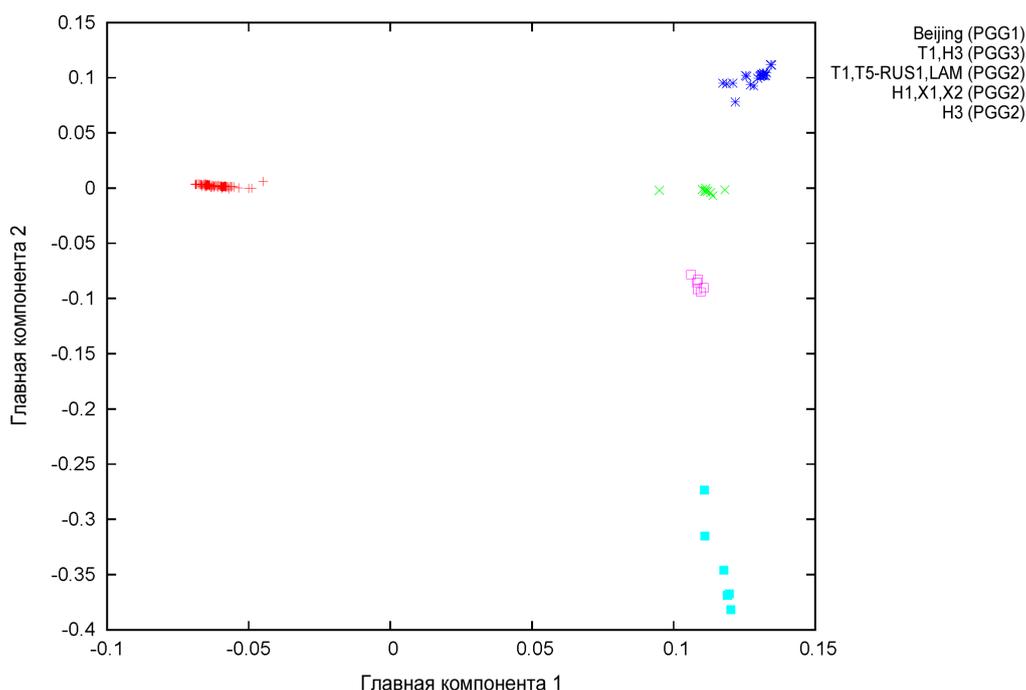


Рис. 2. Расположение популяций *M. tuberculosis*, представленных в Беларуси, в плоскости двух главных компонент

На полном наборе данных, состоящем из 132 организмов и 8000 геномных маркеров, критерий отбора по правилу Кайзера оставляет 28 главных компонент в качестве значимых.

### 2.3. Одномаркерные методы анализа для поиска ассоциированных мутаций

Методы одномаркерного анализа, как правило, являются модификациями классических тестов статистической проверки гипотез и используются для проверки существования зависимостей между отдельными мутациями и изменением фенотипа. В данном случае исследуется статистическая связь между наличием мутации в некоторой позиции генома и развитием лекарственной устойчивости к рассматриваемому препарату. Для этого составляется таблица сопряженности, где в качестве признаков выбираются наименование препарата ( $R$  – устойчивость к этому препарату,  $S$  – чувствительность к нему) и мутация ( $Y$  – присутствие,  $N$  – отсутствие). Подобные таблицы строятся независимо друг от друга для каждой комбинации, состоящей из наименования препарата и мутации. Если рассматриваемая позиция находится в пределах некоторого гена, то при наличии мутации также говорят о присутствии альтернативной аллели гена, а в случае ее отсутствия – преобладающей аллели.

Предположим, что исследуемая группа состоит из  $n_{0*}$  лекарственно-устойчивых бактерий и  $n_{1*}$  лекарственно-чувствительных (табл. 1). Задача статистической проверки гипотез заключается в формулировании основной нулевой гипотезы  $H_0$  о независимости признаков и последующей попытке ее опровержения в пользу альтернативной гипотезы  $H_1$ .

Таблица 1

Таблица сопряженности, рассматриваемая в одномаркерных тестах поиска ассоциаций

Чувствительность к препарату	Мутация присутствует ( $Y$ )	Мутация отсутствует ( $N$ )	Всего
Устойчив к препарату ( $R$ )	$n_{00}$	$n_{01}$	$n_{0*}$
Чувствителен к препарату ( $S$ )	$n_{10}$	$n_{11}$	$n_{1*}$
Всего	$n_{*0}$	$n_{*1}$	$n_{**}$

Для проверки статистических гипотез применялись следующие тесты: критерий Кохрана – Мантеля – Хензеля (Cochran – Mantel – Haenszel, CMH) [3] для анализа таблиц сопряженности, построенных по подгруппам организмов, и  $\chi^2$ -критерий Пирсона с поправкой EIGENSTRAT [4] для учета структуры популяции на основании рассчитанных в подразд. 2.2 значимых главных компонент. Необходимость внесения поправок по отношению к популяционной структуре обусловлена тем, что варьирование частоты встречаемости некоторых мутаций в исследуемых организмах может объясняться их принадлежностью к различным популяциям, а не воздействием лекарственных препаратов.

Статистический тест Кохрана – Мантеля – Хензеля использовался для анализа таблиц сопряженности размерности  $2 \times 2 \times K$ , где  $K$  означает число популяций. Этот тест позволяет нивелировать влияние фактора стратификации на конечный результат. Нулевая гипотеза  $H_0$  в этом случае заключается в предположении об условной независимости двух признаков (наличие лекарственной устойчивости, тип аллели) в случае принадлежности исследуемых организмов к одной из групп. Вероятностная модель при нулевой гипотезе предполагает, что маргинальные итоги фиксированы для каждой из  $K$  групп организмов. Тогда число наблюдений  $n_{00(k)}$  в таблице с номером  $k$  описывается гипергеометрическим распределением вероятностей

тей  $n_{00(k)} \sim HG(n_{0*(k)}, n_{*(k)}, n_{*0(k)}) = \frac{\binom{n_{0*(k)}}{n_{00(k)}} \binom{n_{*(k)}}{n_{10(k)}}}{\binom{n_{***(k)}}{n_{*0(k)}}}$ , где  $\binom{n}{p} = \frac{n!}{p!(n-p)!}$  обозначает би-

номиальный коэффициент. Отсюда можно рассчитать значения и в остальных ячейках таблицы. Если учесть, что математическое ожидание и дисперсия гипергеометрического распределения вероятностей имеют вид

$$E\{n_{00(k)}\} = \frac{n_{0*(k)} n_{*0(k)}}{n_{***(k)}} \quad \text{и} \quad Var\{n_{00(k)}\} = \frac{n_{0*(k)} n_{1*(k)} n_{*0(k)} n_{*1(k)}}{n_{***(k)}^2 (n_{***(k)} - 1)}$$

соответственно, то статистика Кохрана – Мантеля – Хензеля задается формулой

$$CMH = \frac{\left| \sum_{k=1}^K (n_{00(k)} - E\{n_{00(k)}\}) \right|^2}{\sum_{k=1}^K Var\{n_{00(k)}\}} \quad (1)$$

и имеет распределение  $\chi_1^2$  при  $n \rightarrow \infty$  [5].

Выполнение гипотезы  $H_0$  также эквивалентно равенству единице отношений шансов

$$R_{(k)} = \frac{n_{00(k)}/n_{10(k)}}{n_{01(k)}/n_{11(k)}}, \quad k = \overline{1, K}, \quad \text{в каждой из } K \text{ подгрупп, т. е. } H_0: R_{(1)} = R_{(2)} = \dots = R_{(k)} = 1. \text{ Следует}$$

отметить, что тест валиден, когда отношения шансов имеют одинаковые направления и масштабы, и не является эффективным для случая, когда существуют подгруппы  $K_1$  и  $K_2$ , такие, что для любого элемента  $k_1 \in K_1$  отношение шансов  $R_{(k_1)} > 1$  и в то же время для любого элемента  $k_2 \in K_2$  отношение шансов  $R_{(k_2)} < 1$ . В численных экспериментах на предоставленных данных использовалась версия теста, реализованная в пакете PLINK [6].

Еще одним способом учесть зависимости между наблюдениями ввиду генетического сходства соответствующих им организмов, принадлежащих одной популяционной группе, является внесение поправок в данные, которые поступают на вход классических статистических тестов. В частности, использовался классический  $\chi^2$ -критерий Пирсона с поправкой на популяционную структуру, реализованный в пакете EIGENSTRAT [4]. Эта поправка представляет собой модификацию матрицы генотипов  $X$ , по которой затем вычисляется тестовая статистика. Пусть  $x_{ij}$ ,  $i = \overline{1, n}$ ,  $j = \overline{1, m}$  – элемент исходной матрицы  $X$ , где  $i$  – номер организма,  $j$  – позиция мутации. Будем считать, что ее элементы были предварительно нормированы описан-

ным в подразд. 2.2 способом. Пусть ее сингулярное разложение имеет вид  $X = USV^T$ . Тогда для каждой  $l$ -й оси вариаций элементы скорректированной матрицы генотипов рассчитываются по формуле  $x_{ij}^{adj} = x_{ij} - \gamma_j u_{li}$ , где  $\gamma_j = \frac{\sum_{i=1}^n u_{li} x_{ij}}{\sum_{i=1}^n u_{li}^2} = \sum_{i=1}^n u_{li} x_{ij}$ . В матричном виде эти соотношения записываются как  $X_l^{adj} = X - Xu_l$ , т. е. из данных вычитается проекция на  $l$ -ю главную компоненту. Предполагается, что данные были предварительно центрированы относительно начала координат, а величина проекции расстояния между любыми двумя наблюдениями вдоль выбранной главной компоненты пропорциональна генетическому расстоянию между соответствующими организмами. Таким образом устраняются значительные вариации, вызванные принадлежностью организмов к различным популяциям. В силу того что оси главных компонент ортогональны, эта поправка может быть вычислена независимо для каждой оси. Поправка для вектора фенотипов рассчитывается как  $y_{(l)}^{adj} = y_l - u_l$ .

В результате статистической проверки гипотез во всех представленных тестах вычисляется значение некоторой статистики, имеющей известное распределение, и рассчитывается  $p$ -значение. Общепринятым значением ошибки первого рода в полногеномных тестах ассоциаций считается  $\alpha = 5 \times 10^{-8}$  [7]. В случае применения стандартного порога  $\alpha = 0,01$  для ошибки первого рода вероятность сделать хотя бы одну ошибку уже при 100 тестах составляет  $1 - (1 - 10^{-2})^{100} \approx 0,63$ , а при 1000 тестах –  $1 - (1 - 10^{-2})^{1000} \approx 1$ . Для уменьшения вероятности ошибок при проведении множественной проверки гипотез используются методы поправки  $p$ -значений: поправки Бонферони, Бонферони – Холма [8] и Бенджамини – Хохберга [9].

При представлении результатов в работе использовалась поправка Бенджамини – Хохберга. Она гарантирует, что ожидаемая доля ложных отклонений гипотез (false discovery rate, FDR) не будет превосходить  $\alpha$ . Алгоритм проверки гипотез по данному методу работает итерационно с упорядоченным по возрастанию набором  $p$ -значений, отвергая все гипотезы, для которых  $p_{(j)} < \alpha \cdot j/n$ .

#### 2.4. Многомаркерные методы анализа для поиска ассоциированных мутаций

Несмотря на применяемые поправки, большим недостатком одномаркерных методов является то, что они не учитывают парные взаимодействия и взаимодействия более высоких порядков между исследуемыми генетическими вариациями. В то же время существуют методы, именуемые в литературе многомаркерными методами анализа ассоциаций, которые могут реалистично моделировать влияние множества генетических маркеров одновременно. Однако применение многомаркерного анализа сопряжено с трудностями, возникающими при решении задач большой размерности, когда число исследуемых мутаций значительно превосходит доступное число наблюдений. Чтобы сгладить влияние этих факторов, применялись методы отбора признаков и способы регуляризации классических методов машинного обучения.

##### 2.4.1. Методы на основе регуляризованной логистической регрессии

Логистическая регрессия с регуляризацией вектора параметров по  $l_1$ -норме (метод Лассо) в выполненных экспериментах показала лучшие результаты по сравнению с регуляризацией по  $l_2$ -норме (гребневая регрессия). Предпочтительным вариантом стало применение метода Elastic Net (компромиссный вариант между регуляризацией по  $l_1$ - и  $l_2$ -нормам). В работе [10] доказывается, что при регуляризации Лассо для корректной оценки параметров достаточно иметь выборку, число наблюдений в которой оценивается как логарифм от числа параметров модели. Формально модель логистической регрессии задается соотношением

$$p(x) = \frac{1}{1 + e^{-(\beta_0 + x\beta^T)}},$$

где  $p(x)$  соответствует вероятности принадлежности генотипа  $x$  к классу лекарственно-устойчивых бактерий,  $\beta$  – вектор коэффициентов.

Обозначим через  $l(x_i, y_i, \theta)$  функцию потерь при классификации с помощью логистической регрессии  $i$ -го организма с вектором генотипа  $x_i$ , значением фенотипа  $y_i$  и вектором параметров  $\theta = (\beta_0, \beta)$  размерности  $m+1$ . Тогда для поиска оценок вектора параметров минимизируется функционал

$$\sum_{i=1}^n l(x_i, y_i, \theta) + \lambda_1 \|\theta\|_{l_1} + \lambda_2 \|\theta\|_{l_2} = \sum_{i=1}^n l(x_i, y_i, \beta) + \lambda \left( \alpha \sum_{j=0}^m |\beta_j| + (1-\alpha) \sum_{j=0}^m \beta_j^2 \right) \rightarrow \min_{\beta}.$$

Для решения данной оптимизационной задачи использовался метод циклического покоординатного спуска [11], реализованный в библиотеке LARS пакета R. Несмотря на то что целевая функция при такой регуляризации не является дифференцируемой из-за недифференцируемости модуля, она обладает полезным свойством разреженности, т. е. обращением в нуль коэффициентов при увеличении параметра  $\lambda_1$ . Для признания коэффициентов значимыми при некотором  $\lambda$  и  $\alpha \approx 0,7$ , которые были выбраны как наиболее предпочтительные при обучении на имеющихся наборах данных, использовалось ограничение на малость коэффициентов, т. е. значимыми признавались мутации из подмножества  $S$ , такие, что для любого  $s \in S$  абсолютное значение коэффициента  $|\beta_s| \geq \tau$ . Одним из главных недостатков подобной регуляризации является включение в итоговый результат шумовых признаков до того, как были включены все значимые признаки, и ошибки при отборе коррелирующих признаков. Подход с использованием сетей релевантности признаков, описанный в подразд. 2.5, позволяет частично корректировать найденные решения и получать более точный результат даже в случае переобучения.

#### 2.4.2. Линейная смешанная модель

Линейная смешанная модель в статистической генетике применяется для анализа геномных последовательностей, для которых характерна сложная парадигма наследственности и изменчивости. Наиболее широкое распространение в задачах вычислительной биологии получила так называемая смешанная линейная модель, описанная Хендерсоном [12]. В общем виде смешанная модель может быть представлена в матричной форме:

$$y = X\beta + u + \varepsilon, \quad (2)$$

где  $y$  – вектор фенотипов;  $X$  – матрица генотипов;  $\beta$  – вектор параметров, определяющих значимости мутаций для развития лекарственной устойчивости;  $u$  – вектор случайных эффектов, имеющий многомерное нормальное распределение  $u \sim N_n(0, \lambda \sigma_\varepsilon^2 K)$ . Здесь  $\lambda = \sigma_u^2 / \sigma_\varepsilon^2$  – отношение между дисперсиями, обусловленными внутренними (генетическими) и внешними факторами,  $\sigma_\varepsilon^2$  – дисперсия случайной ошибки, а  $K$  – матрица родства организмов (kinship matrix) размерности  $n \times n$ , где  $n$  – число исследуемых организмов. Таким образом, благодаря добавлению случайных эффектов модель позволяет учитывать следствия от наличия популяционной структуры. Одним из способов задания матрицы родства является оценка  $K = XX^T / m$ , где  $m$  – число признаков (генетических маркеров). Как правило, матрица  $K$  центрируется так, чтобы среднее значение в каждом столбце равнялось нулю. Вектор случайных ошибок  $\varepsilon$ , называемый также вектором эффектов окружающей среды, имеет многомерное нормальное распределение вида  $N_n(0, \sigma_\varepsilon^2 I)$ , где  $I$  – единичная матрица соответствующей размерности.

Вычисление точечных оценок параметров  $\beta$ ,  $\sigma_\varepsilon^2$  и  $\lambda$  выполняется по методу максимального правдоподобия [13]. Функция правдоподобия для линейной смешанной модели (2) имеет вид  $l(y; \lambda, \sigma_\varepsilon^2, \beta) = -\frac{1}{2} \left( n \log(2\pi\sigma_\varepsilon^2) + \log|H| + \frac{1}{\sigma_\varepsilon^2} (y - X\beta)^T H^{-1} (y - X\beta) \right)$ , где  $H = \lambda K + I$ . Если считать параметр  $\lambda$  известным, то максимум функции достигается при оценках параметров

$\hat{\beta} = (X^T H^{-1} X)^{-1} X^T H^{-1} \mathbf{y}$  и  $\hat{\sigma}_\varepsilon^2 = (y - X\hat{\beta})^T H^{-1} (y - X\hat{\beta}) / n$ . Подставляя эти значения в выражение для функции правдоподобия, получим, что окончательное вычисление оценки максимума правдоподобия сводится к оптимизации по параметру  $\lambda$  функционала

$$l(\mathbf{y}; \lambda) = -\frac{1}{2} \left( n \log \left( \frac{2\pi (y - X\beta)^T H^{-1} (y - X\beta)}{n} \right) + \log |H| + n \right).$$

Для проверки ограничений на параметры модели используется один из классических статистических тестов: тест отношения правдоподобия, тест Вальда или тест множителей Лагранжа. При оценивании значимости отдельных генетических маркеров с индексами  $j = \overline{1, m}$  по выборочным данным проверяется нулевая гипотеза  $H_0: \beta_j = 0$  против альтернативы  $H_1: \beta_j \neq 0$ . В частности, в тесте отношения правдоподобия сравниваются оценки максимума функции правдоподобия, полученные в нулевой и альтернативной моделях. Построенная на их основе тестовая статистика  $z_{LR} = 2(l_1(\mathbf{y}; \hat{\lambda}_1) - l_0(\mathbf{y}; \hat{\lambda}_0))$  асимптотически имеет распределение  $\chi_1^2$ , если верна нулевая гипотеза.

На практике представляет интерес оценка степени наследуемости признака (доля дисперсии фенотипа, объясненная уравнением регрессии в нулевой модели, т. е. при  $\beta = 0$ ). Значение данной величины определяется по формуле  $PVE = \sigma_u^2 / (\sigma_u^2 + \sigma_\varepsilon^2) = \lambda / (\lambda + 1)$  и служит аппроксимацией для оценки вклада генетических факторов в изменение фенотипа. В случае когда величина наследуемости близка к единице, изменчивость наблюдаемого фенотипа (например, устойчивость/чувствительность к исследуемому лекарственному препарату) может быть полностью объяснена унаследованными генетическими факторами. Если же величина наследуемости близка к нулю, изменчивость фенотипа в большей степени определяется факторами внешней среды.

#### 2.4.3. Методы стохастического отбора значимых признаков

Алгоритм стохастического поиска важных признаков основан на байесовской процедуре отбора переменных. Он идентифицирует комбинации полиморфизмов (мутаций), которые обладают наилучшей прогностической способностью, выполняя поиск среди моделей линейной регрессии с числом факторов, не превосходящим некоторой величины  $k$ , либо иерархических лог-линейных моделей с аналогичным ограничением по числу переменных. Так как ожидаемое число значимых мутаций не является большим, в работе использовались значения  $k \in [2, 5]$ .

Согласно принципу Байеса процедура выбора оптимальной модели состоит в поиске варианта, имеющего максимальную апостериорную вероятность среди моделей рассматриваемого класса. Пусть рассматриваются альтернативные модели  $M_1, \dots, M_L$ , образующие множество  $M$ , и заданы их априорные вероятности  $P(M_i)$ . Каждой модели  $M_i$  соответствует определенная плотность распределения  $p_i(\theta)$  и условная плотность  $p_i(X, \mathbf{y} | \theta)$ , где  $\theta \in \Theta_i$  – вектор параметров из множества настраиваемых параметров модели  $M_i$ ;  $X$  – матрица генотипов;  $\mathbf{y}$  – вектор фенотипов. Тогда апостериорные вероятности моделей  $M_i$  при известных наблюдениях  $X, \mathbf{y}$  вычисляются с помощью формулы Байеса  $P(M_i | X, \mathbf{y}) = P(X, \mathbf{y} | M_i) P(M_i) / \sum_{j=1}^L P(X, \mathbf{y} | M_j) P(M_j)$  и для поиска наилучшей модели требуется максимизировать по  $\mu \in M$  рассчитанную таким образом функцию апостериорной вероятности  $P(\mu | X, \mathbf{y})$ .

Каким бы ни было выбрано множество моделей  $M$ , в котором осуществляется поиск, на нем должна быть определена функция  $nbh(\mu)$  для всех  $\mu \in M$ , возвращающая окружение модели  $\mu$ , т. е. множество предшествующих и последующих моделей. Любые две модели из множества  $M$ , например  $\mu$  и  $\mu'$ , должны быть соединены как минимум одной простой цепью  $\mu = \mu_1, \mu_2, \dots, \mu_l = \mu'$ , такой, что  $\mu_j \in nbh(\mu_{j-1})$ ,  $j = \overline{2, l}$ .

Алгоритм, непосредственно выполняющий поиск наилучшей модели, известен в литературе как ориентированный на моду стохастический поиск (mode-oriented stochastic search, MOSS) [14], который формирует множество моделей  $M(c)$  вида

$$M(c) = \left\{ \mu \in M: P(\mu | X, y) \geq c \cdot \max_{\mu' \in M} P(\mu' | X, y) \right\},$$

где  $c \in (0, 1)$ . В процессе своей работы алгоритм поддерживает множество  $S$  текущих моделей, обновляемое по мере поиска. Под  $S(c)$  будем понимать подмножество множества  $S$ , введенное аналогично тому, как вводится  $M(c)$  относительно  $M$ . Пусть помимо  $c$  задана константа  $c'$ , такая, что  $0 < c' < c$  и  $S(c) \subset S(c')$ , и предопределена вероятность  $q$  удаления моделей из множества  $S \setminus S(c)$ . Некоторую модель  $\mu$  будем называть исследованной, если было просмотрено все ее окружение  $\mu' \in nbh(\mu)$ . Таким образом, множество  $S$  содержит исследованные и неисследованные модели. Алгоритм MOSS работает следующим образом:

1. Множество  $S$  произвольным образом инициализируется моделями из множества  $M$ . Для каждой модели  $\mu \in S$  вычисляется и запоминается ее апостериорная вероятность  $P(\mu | X, y)$ , а  $\mu$  помечается как неисследованная.

2. Пусть  $\Lambda \subset S$  – подмножество неисследованных моделей из  $S$ . Выберем некоторую модель  $\mu \in \Lambda$  случайным образом с вероятностью, пропорциональной апостериорной вероятности  $P(\mu | X, y)$  модели. Пометим выбранную модель  $\mu$  как исследованную.

3. Для всех моделей  $\mu' \in nbh(\mu)$  проверяем их принадлежность к  $S$ . Если  $\mu' \notin S$ , то вычисляем и запоминаем ее апостериорную вероятность  $P(\mu' | X, y)$ . Если при этом  $\mu' \in S(c')$ , то добавляем  $\mu'$  в  $S$  и помечаем ее как неисследованную. Если теперь  $\mu'$  имеет наибольшую апостериорную вероятность среди всех моделей в  $S$ , то из  $S$  удаляются модели  $S \setminus S(c)$ .

4. С вероятностью  $q$  из  $S$  удаляются все модели  $S \setminus S(c)$ .

5. Если все модели в  $S$  оказываются исследованными, то из  $S$  удаляются модели  $S \setminus S(c)$  и множество  $S(c)$  является ответом. Иначе возвращаемся к п. 2 алгоритма.

На примере алгоритма MOSS видно, что модели с низкой апостериорной вероятностью автоматически удаляются из рассмотрения, а приоритет на каждой итерации поиска отдается исследованию наиболее перспективных моделей среди просмотренных ранее, что позволяет алгоритму быстро двигаться в сторону моделей с высокой апостериорной вероятностью. Значения гиперпараметров  $c$ ,  $c'$  и  $q$  определяют размер множества просматриваемых моделей, что позволяет ограничивать их перебор.

В отличие от регрессионного подхода к оценке значимости индивидуальных мутаций алгоритм MOSS позволяет ранжировать и сравнивать модели в соответствии с их апостериорной вероятностью. Кроме того, апостериорные вероятности моделей позволяют оценивать значимость отдельных мутаций с помощью байесовского усреднения [15].

### 2.5. Использование графической модели для отбора значимых признаков

Важно отметить, что описанные выше методы, за исключением в какой-то степени алгоритма MOSS, не учитывают взаимозависимости между признаками исследуемой модели. В реальных же данных генетические маркеры достаточно часто коррелируют друг с другом

с коэффициентом корреляции Пирсона, превосходящим по модулю 0,8. Одним из способов представления пространства признаков, который учитывает их взаимозависимости, является так называемая сеть релевантности признаков [16].

Рассмотрим задачу построения сети релевантности признаков для поиска значимых мутаций в некоторой геномной последовательности. Допустим, что поставлена задача обучения с учителем по  $m$  признакам с размером обучающей выборки  $n$ . Тогда сеть релевантности признаков называют бинарное марковское случайное поле, заданное с помощью случайного вектора  $\Xi = (\Xi_1, \Xi_2, \dots, \Xi_m) \in \{0, 1\}^m$  и неориентированного графа  $G(V, E)$ . Состояние вершины  $j \in V$  графа  $G(V, E)$  кодируется при помощи бинарной переменной  $\xi_j$ , принимающей значение  $\xi_j = 1$ , если наличие  $j$ -й мутации связано с развитием лекарственной устойчивости, и  $\xi_j = 0$ , если такая связь не установлена. Любые вершины  $i$  и  $j$ , соответствующие попарно коррелированным мутациям, соединены в графе неориентированным ребром  $(i, j) \in E$ . На вершинах графа  $G(V, E)$  введем унарные потенциалы  $\phi(\Xi_j)$ , характеризующие априорную вероятность того, насколько  $j$ -я мутация определяет лекарственную устойчивость к рассматриваемому лекарственному препарату, а на ребрах – парные потенциалы  $\psi(\Xi_i, \Xi_j)$ . Для заданной таким образом сети вероятность некоторой реализации  $\xi = (\xi_1, \xi_2, \dots, \xi_m)$  случайного вектора  $\Xi$  определяется формулой

$$P(\xi) = \frac{1}{Z} \prod_{j=1}^{|V|} \phi_j(\xi_j) \prod_{(i,j) \in E} \psi_{ij}(\xi_i, \xi_j) = \frac{1}{Z} \exp \left( \sum_{j=1}^{|V|} \log \phi_j(\xi_j) + \sum_{(i,j) \in E} \log \psi_{ij}(\xi_i, \xi_j) \right), \quad (3)$$

где  $Z$  – константа нормализации, а  $|V| = m$ . Определим функцию энергии сети  $E(\xi)$  по формуле

$$E(\xi) = \sum_{j=1}^{|V|} E_j(\xi_j) + \sum_{(i,j) \in E} E_{ij}(\xi_i, \xi_j) + E_0, \quad (4)$$

где  $E_j(\xi_j) = -\log \phi(\xi_j)$ ,  $E_{ij}(\xi_i, \xi_j) = -\log \psi(\xi_i, \xi_j)$ ,  $E_0$  – константа. Тогда конечная цель заключается в поиске конфигурации  $\xi_{opt} = (\xi_1, \xi_2, \dots, \xi_m)$ , которая минимизирует функцию энергии  $E(\xi)$  в формуле (4), что эквивалентно максимизации вероятности  $P(\xi)$  в формуле (3). Скрининг переменных с помощью сетей релевантности состоит из двух этапов: построения сети и статистического вывода в ней.

На этапе построения сети требуется задать функции  $\phi$  и  $\psi$ . Чтобы задача имела решение за полиномиальное время, выбор функции  $\psi$  выполняется с учетом условия субмодулярности [16]:

$$E_{ij}(0, 0) + E_{ij}(1, 1) \leq E_{ij}(0, 1) + E_{ij}(1, 0), \quad (5)$$

которое является необходимым и достаточным для сведения задачи минимизации функции энергии (4) к задаче поиска минимального разреза в соответствующем ей графе  $G'(V', E')$  (способ построения этого графа приведен в конце раздела).

Для обеспечения условия (5) в рассматриваемой задаче можно положить  $\psi(\Xi_i, \Xi_j) = \exp\{\lambda \cdot |r_{ij}| \cdot I(\Xi_i = \Xi_j)\}$ , где  $\lambda > 0$  – положительный параметр;  $r_{ij}$  – коэффициент корреляции признаков;  $I(\Xi_i = \Xi_j)$  – индикаторная переменная, равная единице, если значения

признаков  $\Xi_i$  и  $\Xi_j$  совпадают. Заметим, что такой выбор функции парных потенциалов поощряет ситуации, когда хорошо коррелированные признаки должны оказаться в одном и том же состоянии, т. е. коррелирующие друг с другом мутации одновременно либо являются, либо не являются маркерами лекарственной устойчивости к рассматриваемому препарату.

Выбор унарных потенциалов вершин  $\phi$  достаточно произволен и можно положить  $\phi(\Xi_j) = \exp\{|\Xi_j - q_j|\}$ , где  $q_j = 1 - p_j$  – вероятность того, что  $j$ -й признак не является существенным. Предположим, что при определении значимости каждого отдельного признака использовались методы статистической проверки гипотез, например методы из подразд. 2.3, 2.4, и для каждого признака было вычислено значение некоторой статистики. Пусть  $T$  – вектор этих статистик  $T = (T_1, \dots, T_m)$ , где все  $T_j$  независимы. Тогда каждое  $p_j$  определяется исходя из значений тестовой статистики  $T_j$ . Например,  $p_j = 1$  при  $|T_j| \geq \tau$  и  $p_j = 0$  в противном случае.

Вторым шагом при определении значимости признаков является вероятностный вывод. Для построенной сети релевантности признаков требуется найти вектор  $\xi_{opt}$ , максимизирующий апостериорную вероятность (3). Подставив в формулу (4) выражения для потенциалов вершин и ребер, получим эквивалентную формулировку в виде задачи минимизации функции энергии:

$$\sum_{j=1}^{|V|} |\xi_j - p_j| + \lambda \sum_{i,j=1}^{|V|} I(\xi_i \neq \xi_j) |r_{ij}| \rightarrow \min_{\xi} . \quad (6)$$

Поскольку парные потенциалы  $\psi(\Xi_i, \Xi_j)$  подобраны так, что полученные на их основе функции  $E_{ij}(\Xi_i, \Xi_j) = \lambda I(\Xi_i \neq \Xi_j) |r_{ij}|$  удовлетворяют условиям субмодулярности (5), наиболее вероятная реализация случайного вектора  $\Xi$  может быть найдена с помощью алгоритмов поиска минимального  $s$ - $t$ -разреза в специальном графе  $G'(V', E')$ , который получается из графа  $G(V, E)$  следующим образом. Множество вершин  $V' = V \cup \{s, t\}$  нового графа формируется путем введения двух дополнительных терминальных вершин  $s, t$ . Для построения множества дуг  $E'$  каждая нетерминальная вершина  $j \in V$  соединяется с истоком  $s$  ребром веса  $c_{sj} = |1 - p_j| - |0 - p_j|$ , если  $|1 - p_j| - |0 - p_j| > 0$ , либо со стоком  $t$  ребром веса  $c_{jt} = |0 - p_j| - |1 - p_j|$  в противном случае. Каждая пара нетерминальных вершин  $i, j \in V$  соединяется ребром, вес которого определяется как  $c_{ij} = E_{ij}(0, 1) + E_{ij}(1, 0) - E_{ij}(0, 0) - E_{ij}(1, 1)$ . Отметим, что благодаря ограничениям (5) и выбору функций потенциалов веса дуг в графе  $G'(V', E')$  являются неотрицательными. Любой  $s$ - $t$ -разрез такого графа соответствует разбиению множества вершин  $V'$  на подмножества истока и стока. После нахождения минимального  $s$ - $t$ -разреза в графе  $G'(V', E')$  значения переменных  $\xi_j$  определяются следующим образом:  $\xi_j = 1$ , если вершина  $j$  остается связанной со стоком  $t$  в полученном разрезе, и  $\xi_j = 0$  в противном случае. Величина разреза, рассчитываемая как сумма емкостей всех его ребер, совпадает со значением функционала энергии (4) при таких значениях переменных  $\xi$ .

### 3. Результаты исследования

По итогам анализа имеющихся данных полногеномного секвенирования *M. tuberculosis* и медицинских карт пациентов были отобраны 132 генома и сформированы тестовые наборы данных (табл. 2). Материал из каждого набора был проанализирован с использованием описанной выше методологии.

Таблица 2

Характеристики наборов данных, сформированных на основе результатов лабораторных тестов на чувствительность к лекарственным препаратам для проведения вычислительного эксперимента

Набор данных	Число наблюдений	Условия отбора	Лекарственные препараты, к которым выполняется поиск маркеров лекарственной устойчивости
1	132	–	Все первой линии и офлоксацин
2	48	Чувствителен к аминогликозидам	Офлоксацин
3	54	Чувствителен к офлоксацину	Аминогликозиды
4	23	Чувствителен к аминогликозидам, но устойчив к рифампицину	Офлоксацин
5	63	Чувствителен к амикацину	Капреомицин
6	122	–	Аминогликозиды
7	29	Чувствителен к офлоксацину, но устойчив к рифампицину	
8	48	Чувствителен к аминогликозидам	Все второй линии кроме аминогликозидов
9	122	–	

Для сопоставления результатов отбора значимых мутаций с помощью предложенной методологии с уже известными генетическими маркерами лекарственной устойчивости была использована публичная база данных TBdreamDB [17], которая содержит информацию о большинстве описанных в литературе мутаций к препаратам первого ряда, а также списки полиморфизмов, которые заложены в применяемых на практике тест-системах GenoType MTBDRplusV2 и GenoType MTBDRsl [18, 19].

Для визуализации результатов полногеномных тестов ассоциаций применяют так называемые Манхэттен-графики, в которых ось  $x$  соответствует позициям однонуклеотидных полиморфизмов в геноме, а ось  $y$  – отрицательному десятичному логарифму  $p$ -значений (рис. 3). Чем выше статистическая значимость отдельной мутации, тем больше значение координаты  $y$  соответствующей точки на графике.

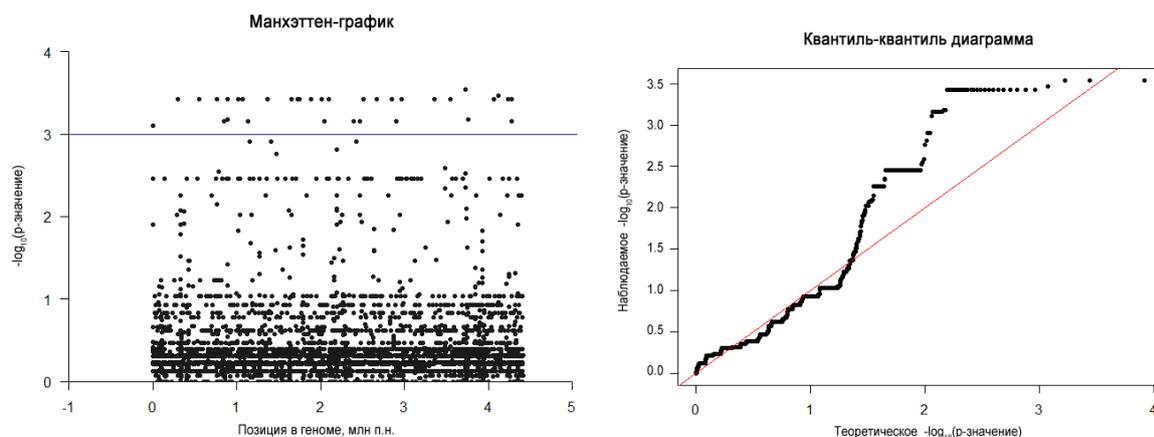


Рис. 3. Манхэттен-график и соответствующая квантиль-квантиль-диаграмма, демонстрирующие результаты применения теста Кохрана – Мантеля – Хензеля для поиска мутаций лекарственной устойчивости к офлоксацину. Для поправки  $p$ -значений применялся метод геномной коррекции [20]

Методы одномаркерного анализа ассоциаций, описанные в разд. 2.3, позволили идентифицировать в качестве валидных для определения лекарственной устойчивости как уже известные мутации, так и ряд новых потенциально значимых полиморфизмов. Для контроля качества результатов одномаркерных методов использовались квантиль-квантиль-диаграммы (Q-Q plots) распределений  $p$ -значений, рассчитанных с учетом поправок на множественные сравнения.

Совокупно по результатам применения смешанных линейных моделей в каждом из экспериментов были найдены множества значимых мутаций и получены оценки параметров модели. Статистическая значимость коэффициентов  $\beta$  оценивалась с помощью теста отношения правдоподобия. При обучении модели применялся скользящий контроль по пяти блокам (5-fold cross-validation) и вычислялись метрики, общепринятые для оценки качества классификации: точность (precision), полнота (recall), F-мера (среднее гармоническое между точностью и полнотой), правильность (accuracy). Истинные значения классов для каждого наблюдения (генома) заимствуются из результатов лабораторных тестов на чувствительность к препарату (табл. 3).

Таблица 3

Результаты классификации для тривиального решающего правила  $\alpha(X, \beta) = \text{sign}(X\beta)$ , где вектор параметров  $\beta$  оценивался с помощью линейной смешанной модели. Значения  $\alpha(X, \beta) = 1$  соответствуют классу  $y = 1$  (устойчивость), значения  $\alpha(X, \beta) = -1$  – классу  $y = 0$  (чувствительность)

Набор данных	Препарат	1	0	Значение истинное/предсказанное				Точность	Полнота	F1	Правильность
				1/1	1/0	0/1	0/0				
1	OFLO	69	63	0,523	0,000	0,417	0,061	0,557	1,000	0,715	0,583
	EMB	102	30	0,083	0,689	0,000	0,227	1,000	0,108	0,195	0,311
	INH	106	26	0,803	0,000	0,000	0,197	1,000	1,000	1,000	1,000
	PZA	28	6	0,824	0,000	0,029	0,147	0,966	1,000	0,983	0,971
	RIF	106	26	0,803	0,000	0,000	0,197	1,000	1,000	1,000	1,000
	STM	110	22	0,826	0,008	0,000	0,167	1,000	0,991	0,995	0,992
2	OFLO	10	38	0,208	0,000	0,000	0,792	1,000	1,000	1,000	1,000
3	AMIK	8	46	0,148	0,000	0,482	0,370	0,235	1,000	0,381	0,519
	CAPR	10	40	0,200	0,000	0,660	0,140	0,233	1,000	0,377	0,340
4	OFLO	10	13	0,435	0,000	0,000	0,565	1,000	1,000	1,000	1,000
6	AMIK	59	63	0,484	0,000	0,516	0,000	0,484	1,000	0,652	0,484
	CAPR	66	51	0,564	0,000	0,436	0,000	0,564	1,000	0,721	0,564
9	CYCL	46	70	0,397	0,000	0,603	0,000	0,397	1,000	0,568	0,397
	ETH	33	89	0,271	0,000	0,631	0,098	0,300	1,000	0,462	0,369
	PARA	22	100	0,115	0,066	0,066	0,754	0,636	0,636	0,636	0,869

Подробнее поясним содержащуюся в табл. 3 информацию. Для примера рассмотрим первую строку этой таблицы. В столбце «Набор данных» приведен номер набора, из которого взяты данные для анализа. В данном случае согласно условиям отбора анализировались все имеющиеся последовательности (см. табл. 2). Название препарата OFLO говорит о том, что выполнялся поиск геномных маркеров лекарственной устойчивости к офлоксацину. Следующие два столбца информируют о составе обучающей выборки: присутствовало 69 образцов, устойчивых к офлоксацину, и 63 чувствительных к нему. Оставшиеся столбцы характеризуют качество предсказаний по набору выявленных методом значимых мутаций: 0,523 – доля образцов в общей выборке, которые являются устойчивыми к офлоксацину и были предсказаны как устойчивые; 0,000 – доля образцов, которые являются устойчивыми, но были отнесены к классу чувствительных; 0,417 – доля образцов, которые являются чувствительными, но были предсказаны как устойчивые; 0,061 – доля образцов, которые являются чувствительными и были отнесены к чувствительным; 0,557 – доля истинно устойчивых среди предсказанных как устойчивые; 1,000 – доля предсказанных как устойчивые среди истинно устойчивых; 0,715 – среднее гармоническое по двум предыдущим показателям; 0,583 – доля правильных предсказаний по всей выборке.

По итогам работы алгоритма стохастического отбора признаков MOSS были получены вероятностные оценки значимости не только отдельных мутаций, но и содержащих их моделей. Такой подход позволяет начать отбор значимых признаков с ранжирования моделей. Несмотря на простоту моделей, среди которых осуществляется перебор вариантов, здесь в них учитываются и возможные взаимодействия между рассматриваемыми признаками. Поскольку общее число признаков после фильтрации данных оставалось достаточно большим (порядка 770 полиморфизмов), была использована двухэтапная схема поиска наилучшей модели. На первом этапе проводился перебор моделей на множестве линейных регрессий с числом регрессоров, не превосходящим трех. На втором этапе взятые признаки послужили построению наиболее вероятной лог-линейной модели, учитывающей взаимодействия между генетическими маркерами. Однако фаза отбора признаков, реализованная в библиотеке genMOSS, в случае большого числа моделей с примерно равной апостериорной вероятностью, эквивалентна полному перебору. Чтобы сократить время вычислений, для формирования входного набора признаков использовались результаты линейной смешанной модели и регуляризованной логистической регрессии. Множества найденных с их помощью мутаций, учитывая поправки на корреляционную структуру признаков, значительно меньше исходного множества признаков. Это позволяет увеличивать число регрессоров в рассматриваемых лог-линейных моделях до пяти-шести. Дальнейшее увеличение сложности моделей приводит к возрастанию числа ложноположительных результатов и усложняет биологическую интерпретацию результатов. При обучении моделей применялись методы скользящего контроля и вычислялись классические метрики оценки качества классификации (табл. 4).

Таблица 4

Результаты классификации с помощью лучших лог-линейных моделей, найденных MOSS

Набор данных	Препарат	1	0	Значение истинное/предсказанное				Точность	Полнота	F1	Правильность
				1/1	1/0	0/1	0/0				
1	OFLO	69	63	0,394	0,130	0,030	0,446	0,929	0,752	0,831	0,840
	EMB	102	30	0,759	0,015	0,030	0,197	0,962	0,981	0,972	0,955
	INH	106	26	0,788	0,015	0,000	0,196	1,000	0,981	0,990	0,985
	PZA	28	6	0,813	0,000	0,000	0,187	1,000	1,000	1,000	1,000
	RIF	106	26	0,698	0,105	0,000	0,197	1,000	0,869	0,930	0,895
	STM	110	22	0,795	0,038	0,000	0,167	1,000	0,954	0,977	0,962
2	OFLO	10	38	0,206	0,000	0,022	0,772	0,902	1,000	0,949	0,978
3	AMIK	8	46	0,103	0,034	0,000	0,863	1,000	0,750	0,857	0,966
	CAPR	10	40	0,160	0,040	0,000	0,800	1,000	0,800	0,889	0,960
4	OFLO	10	13	0,386	0,050	0,029	0,536	0,931	0,885	0,908	0,921
6	AMIK	59	63	0,410	0,074	0,000	0,515	1,000	0,847	0,917	0,926
	CAPR	66	51	0,413	0,152	0,000	0,436	1,000	0,731	0,845	0,848
9	CYCL	46	70	0,180	0,217	0,179	0,424	0,502	0,453	0,477	0,604
	ETH	33	89	0,073	0,198	0,008	0,721	0,905	0,270	0,415	0,794
	PARA	22	100	0,082	0,099	0,000	0,819	0,929	0,454	0,610	0,901

Применение логистической регрессии с регуляризацией показало, что наилучшие результаты классификации достигались при использовании компромиссной регуляризации Elastic Net с коэффициентом  $\alpha = 0,7$  по  $l_1$ -норме. Для отбора переменных было задано ограничение на минимальное значение модуля коэффициентов, признаваемых значимыми, равное  $10^{-3}$ . При обучении модели применялся скользящий контроль по 10 блокам (10-fold cross-validation) и вычислялись классические метрики оценки качества классификации (табл. 5).

Таблица 5

Результаты классификации с помощью логистической регрессии, регуляризованной по методу Elastic Net с параметром  $\alpha = 0,7$

Набор данных	Препарат	1	0	Значение истинное/предсказанное				Точность	Полнота	F1	Правильность
				1/1	1/0	0/1	0/0				
1	OFLO	69	63	0,515	0,008	0,015	0,462	0,971	0,986	0,978	0,977
	EMB	102	30	0,773	0,000	0,008	0,220	0,990	1,000	0,995	0,992
	INH	106	26	0,803	0,000	0,000	0,200	1,000	1,000	1,000	1,000
	RIF	106	26	0,803	0,000	0,000	0,200	1,000	1,000	1,000	1,000
	PZA	28	6	0,824	0,000	0,000	0,177	1,000	1,000	1,000	1,000
	STM	110	22	0,833	0,000	0,000	0,167	1,000	1,000	1,000	1,000
2	OFLO	10	38	0,208	0,000	0,000	0,792	1,000	1,000	1,000	1,000
3	AMIK	8	46	0,111	0,037	0,000	0,852	1,000	0,75	0,857	0,963
	CAPR	10	40	0,200	0,000	0,000	0,800	1,000	1,000	1,000	1,000
4	OFLO	10	13	0,435	0,000	0,000	0,565	1,000	1,000	1,000	1,000
6	AMIK	59	63	0,451	0,033	0,000	0,516	1,000	0,932	0,965	0,967
	CAPR	66	51	0,564	0,000	0,000	0,436	1,000	1,000	1,000	1,000
9	CYCL	46	70	0,388	0,009	0,000	0,603	1,000	0,978	0,989	0,991
	ETH	33	89	0,271	0,000	0,000	0,730	1,000	1,000	1,000	1,000
	PARA	22	100	0,131	0,049	0,000	0,820	1,000	0,727	0,842	0,951

Стоит отметить, что среди рассмотренных методов алгоритм MOSS при достаточно хороших показателях качества классификации выдает наименьшее число значимых признаков, среди которых, как правило, присутствуют уже описанные в литературе мутации лекарственной устойчивости. Смешанные линейные модели и регуляризованная логистическая регрессия идентифицируют гораздо большее количество значимых признаков, что может свидетельствовать о переобучении этих моделей и включении в итоговое множество шумовых признаков. Одномаркерные тесты также оставляют достаточно большое число мутаций в качестве связанных с развитием лекарственной устойчивости.

Анализ значимых признаков, получаемых на выходе описанных выше алгоритмов, показал, что без дополнительных корректировок эти множества не обладают свойством непротиворечивости, т. е. существуют такие пары коррелирующих мутаций, среди которых одна признается значимой, а другая нет. С помощью методов отбора переменных, основанных на применении сетей релевантности признаков (см. подразд. 2.6), удастся сократить множества значимых мутаций как минимум вдвое для большинства тестов за счет удаления потенциально ложноположительных результатов. На рис. 4 показаны графики изменения размеров множества значимых мутаций в зависимости от изменения параметра  $\lambda$  в функционале (6) для теста на лекарственную устойчивость к офлоксацину и соответствующий график изменения  $F$ -меры как единой метрики, характеризующей точность и полноту. В большинстве случаев заметное сокращение множества значимых мутаций приводит к незначительному ухудшению качества классификации.

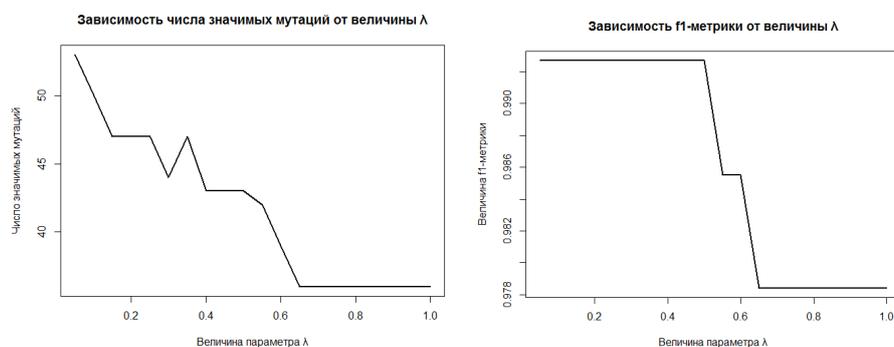


Рис. 4. Графики изменения множества значимых мутаций для лекарственного препарата офлоксацина в результате применения сетей релевантности признаков для корректировки результатов регуляризованной логистической регрессии

При сравнении множества найденных маркеров лекарственной устойчивости с известными ранее были подтверждены многие мутации, входящие в состав широко используемых тест-систем GenoType MTBDRplus/MTBDRsl для препаратов первой линии, а также найдены вероятные мутации, вызывающие резистентность к препаратам второй линии.

### Заключение

Поиск по международным базам данных показал, что исследование микобактерии *M. tuberculosis* является одним из быстро развивающихся направлений. Несмотря на то что в мире неоднократно проводились работы по анализу геномов *M. tuberculosis*, направленные на поиск мутаций лекарственной устойчивости, такие исследования ограничивались применением методов одномаркерного анализа ассоциаций и изучением препаратов первой линии.

В настоящей работе предложена весьма универсальная методология поиска ассоциаций в геномах патогенных микроорганизмов, позволяющая оценивать вклад мутаций в совокупности и корректировать результаты в случае их противоречивости. Всего в рамках исследования, проводимого в Беларуси, было выполнено секвенирование микобактерий, выделенных у 132 пациентов с различными формами туберкулеза легкого. В дополнение к теоретическому анализу методов осуществлена разработка программного обеспечения в среде R, непосредственно предназначенного для анализа геномов.

Авторы выражают благодарность Республиканскому научно-практическому центру пульмонологии и фтизиатрии Министерства здравоохранения Республики Беларусь за содействие и помощь в предоставлении данных.

### Список литературы

1. Borrell, S. Infectiousness, reproductive fitness and evolution of drug-resistant *Mycobacterium tuberculosis* / S. Borrell, S. Gagneux // *Intern. J. Tuberc. Lung Dis.* – 2009. – № 13(12). – P. 1456–1466.
2. Patterson, N. Population structure and eigenanalysis / N. Patterson, A. L. Price, D. Reich // *PLoS Genet.* – 2006. – № 2(12). – P. 2074–2093.
3. Mantel, N. Statistical aspects of the analysis of data from retrospective studies of disease / N. Mantel, W. Haenszel // *J. National Cancer Inst.* – 1959. – № 22(4). – P. 719–748.
4. Principal components analysis corrects for stratification in genome-wide association studies / A.L. Price [et al.] // *Nat. Genet.* – 2006. – № 38(8). – P. 904–909.
5. Agresti, A. An introduction to categorical data analysis / A. Agresti. – Wiley, 2002. – Ch. 6. – P. 231–236.
6. PLINK: a tool set for whole-genome association and population-based linkage analyses / S. Purcell [et al.] // *Am. J. Hum. Genet.* – 2007. – № 81(3). – P. 559–575.
7. Bush, W.S. Genome-wide association studies / W.S. Bush, J.H. Moore // *PLoS. Comput. Biol.* – 2012. – № 8(12). – P. 1–11.
8. Holm, S. A simple sequentially rejective Bonferroni test procedure testing / S. Holm // *Scandinavian Journal of Statistics.* – 1979. – № 6. – P. 65–70.
9. Benjamini, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing / Y. Benjamini, Y. Hochberg // *Journal of the Royal Statistical Society.* – 1995. – № 57. – P. 289–300.
10. Ng, A.Y. Feature selection,  $L_1$  vs.  $L_2$  regularization, and rotational invariance / A.Y. Ng // *Proc. of the 21st Intern. Conf. on Machine Learning.* – Ban, Canada, 2004.
11. Friedman, J. Regularization paths for generalized linear models via coordinate descent / J. Friedman, T. Hastie, R. Tibshirani // *Journal of Statistical Software.* – 2010. – № 33(1). – P. 1–22.
12. Henderson, C.R. Sire evaluation and genetic trends / C.R. Henderson // *Proc. Anim. Breeding and Genetic Symp. in honor of Dr. J.L. Lush.* – Champaign, 1973. – P. 10–41.
13. Zhou, X. Genome-wide efficient mixed-model analysis for association studies / X. Zhou, M. Stephens // *Nature Genetics.* – 2012. – № 44(7). – P. 821–824.

14. Dobra, A. The Mode oriented stochastic search (MOSS) for log-linear models with conjugate priors / A. Dobra, H. Massam // *Statistical Methodology*. – 2010. – № 7. – P. 240–253.
15. Applications of the mode oriented stochastic search (MOSS) algorithm for discrete multi-way data to genomewide studies / A. Dobra [et al.] // *Bayesian Modeling in Bioinformatics*. – CRC Press, 2010. – Ch. 3. – P. 63–94.
16. Kolmogorov, V. What energy functions can be minimized via graph cuts? / V. Kolmogorov, R. Zabih // *IEEE Transactions on Pattern Analysis and Machine Intelligence*. – 2004. – № 26(2). – P. 147–159.
17. Tuberculosis drug resistance mutation database / A. Sandgren [et al.] // *PLoS Med*. – 2009. – № 6(2). – P. 132–136.
18. GenoType MTBDRplus – your test system for a fast and reliable way to detect MDR-TB // *Hain Lifesciences* [Electronic resource]. – 2015. – Mode of access : <http://www.hain-lifescience.de/en/products/microbiology/mycobacteria/genotype-mtbdplus.html>. – Date of access : 10.05.2015.
19. GenoType MTBDRsl – your important assistance for detection of XDR-TB // *Hain Lifesciences* [Electronic resource]. – 2015. – Mode of access : <http://www.hain-lifescience.de/en/products/microbiology/mycobacteria/genotype-mtbdsl.html>. – Date of access : 10.05.2015.
20. Devlin, B. Genomic control for association studies / B. Devlin, K. Roeder // *Biometrics*. – 1999. – № 55. – P. 997–1004.

**Поступила 10.10.2015**

<sup>1</sup>*Объединенный институт проблем информатики НАН Беларуси, Минск, Сурганова, 6  
e-mail: roma.sergeev@gmail.com*

<sup>2</sup>*EPAM Systems, Минск, Академика Купревича, 1/1*

<sup>3</sup>*Office of Cyber Infrastructure and Computational Biology, National Institute of Allergy and Infectious Diseases, NIH, Bethesda, MD, USA*

**R.S. Sergeev, I.S. Kavaliou, A.V. Tuzikov, A. Rosenthal, A. Gabrielian**

**ALGORITHMS FOR IDENTIFYING DRUG-RESISTANCE  
MUTATIONS IN M. TUBERCULOSIS GENOMES**

Analysis of whole-genome sequences often leads to problems of large dimensionality where the number of parameters exceeds the number of available observations. We offer methodology of genome-wide association study and investigate various approaches to assess contribution of mutations in the emergence and development of drug resistance in *Mycobacterium tuberculosis*. We present the results of our experiments aimed at identifying resistance mutations to the major anti-TB drugs based on data obtained from patients in Belarus.

## ЗАЩИТА ИНФОРМАЦИИ

УДК 003.26

Т.В. Галибус, В.В. Краснопрошин

**КОНЦЕПТУАЛЬНОЕ МОДЕЛИРОВАНИЕ И ОРГАНИЗАЦИЯ МЕХАНИЗМОВ ЗАЩИТЫ ИНФОРМАЦИИ В РАСПРЕДЕЛЕННЫХ СИСТЕМАХ**

*Рассматриваются проблемы, связанные с организацией механизмов защиты информации в распределенных компьютерных системах. Предлагается двухуровневая иерархия моделей, которая позволяет упростить выбор эффективной стратегии защиты в зависимости от типа анализируемой системы. Строится модель сетевого механизма безопасности и формулируется методология последовательной настройки механизмов, согласованных с предложенной иерархией моделей.*

**Введение**

В последнее время все большее число прикладных информационных систем (ИС) содержат в себе компоненты распределенной обработки информации [1, 2]. Таковыми, например, являются системы управления содержимым (content management), контроля версий, любые облачные платформы и сервисы и т. д. Современные ИС уже на этапе их проектирования ориентированы на глобальную информационную сеть, а значит, по умолчанию являются распределенными. Ключевые факторы уязвимости таких систем – всевозрастающие открытость и гибкость их инфраструктуры. Поэтому актуальной в настоящее время является проблема построения комплексной, надежной и всеобъемлющей системы защиты информации (СЗИ) [3].

Существует множество механизмов защиты, с помощью которых можно организовать систему безопасности, отвечающую различным целям [4]. Однако универсальной методики, которая позволяла бы снизить затраты на проектирование и автоматизировать процесс построения согласованной с конфигурацией распределенных компьютерных систем (РКС) системы защиты, к сожалению, нет [3, 5].

В работе предлагается подход, основанный на использовании формальных с точки зрения защиты информации моделей, который позволяет описывать различные конфигурации РКС и автоматизировать процесс построения соответствующих систем защиты. Использование формальных моделей дает возможность не только провести классификацию РКС и механизмов их защиты, но и оценить состояние защищенности самой системы.

Классификация общих информационных объектов с точки зрения защиты информации приведена в стандарте СТБ 34.101.30-2007 [6]. Авторами сформулирована специализированная методология проектирования СЗИ, в рамках которой предложена иерархия моделей и механизмов, в наибольшей степени подходящая для организации системы проектирования СЗИ и позволяющая автоматизировать процесс построения СЗИ для РКС в целом.

**1. Общая модель распределенной информационной системы**

Прежде чем сформулировать задачу на построение обобщенной СЗИ и определить методику ее решения, необходимо подробно рассмотреть объект защиты и определить характеристики, важные для построения такой системы.

Распределенные компьютерные системы появились в начале 1990-х гг. Они прошли определенный путь своего развития, при этом парадигма и технология их разработки неоднократно менялись [1].

По мнению многих авторов [2, 5, 7], базовыми составляющими, определяющими характер современных РКС на самом общем уровне, являются архитектура системы (А), сетевые технологии (Т) и тип управления системой (М).

*Архитектура* системы определяет структуру и состав системы, ее функциональное наполнение. Фактически она определяет общую взаимосвязь логической и физической структур системы и описывает организацию работы ее сервисных компонентов.

*Коммуникационные технологии* осуществляют связь между элементами системы и реализуют процесс обмена данными между приложениями на различных компьютерных системах и платформах.

*Управление* организует взаимодействие элементов системы (функционирование процессов внутри системы). Оно определяет степень централизации выполнения административных функций системы.

Принимая во внимание приведенные выше рассуждения и используя методологию построения концептуальных моделей, любую РКС формально можно записать в виде следующего кортежа:

$$C = (A, T, M).$$

При этом каждый компонент такой модели может принимать различные значения, определяя тем самым конфигурацию конкретной РКС. Ниже перечислены известные варианты указанных компонентов:

*A<sub>1</sub>. Клиент-серверная архитектура.* Сервисы (сетевая нагрузка) в таких системах распределены между поставщиками услуг (серверами) и заказчиками услуг (клиентами). В силу высокой централизации компонентов такие системы, как правило, являются хорошо защищенными, но недостаточно гибкими. Кроме того, вероятность отказов в этих системах достаточно велика.

*A<sub>2</sub>. Сервис-ориентированная архитектура (СОА).* В системах с такой архитектурой используются независимые или слабо связанные компоненты, ориентированные на выполнение конкретной функциональности (сервиса). Поставщиками и заказчиками услуг могут быть любые узлы сети, а четко определенный интерфейс сервиса и стандартизированный вызов обеспечивают независимость системы от программно-аппаратной платформы и вызывающего приложения. На сегодняшний день данный вид архитектуры практически является стандартом в корпоративных системах.

*A<sub>3</sub>. Архитектура одноранговых сетей.* В данном случае все узлы сети являются равноправными участниками системы. При этом не существует четкой концепции выполнения конкретной функциональности. Такие системы характеризуются высокой гибкостью, но очень слабым уровнем защищенности.

*A<sub>4</sub>. Архитектура грид.* Такая архитектура представляет собой объединение вычислительных ресурсов нескольких машин для решения трудоемких научных задач и проведения объемных вычислений. Компьютеры в этом случае объединяются в кластеры, и их ресурсы функционируют согласованно с целью выполнения операций, необходимых для решения поставленных задач. В таких системах отсутствует концепция сервиса (услуги). Данная архитектура не совсем подходит для коммерческих организаций и корпоративных сетей.

*A<sub>5</sub>. Облачная архитектура.* Фактически является комбинацией СОА и грид-архитектуры. Иными словами, облачная архитектура представляет собой объединение ресурсов различных машин с целью выполнения конкретных функций. Данное объединение называется виртуализацией, поскольку поставщиком услуг фактически является виртуальный суперкомпьютер. В настоящий момент проблема стандарта безопасности для таких систем является наиболее актуальной как для международных, так и для государственных структур [8].

В настоящее время используются следующие типы коммуникационных технологий [1, 7]:

*T<sub>1</sub>. Объектная технология (CORBA, DCOM).* Технологический стандарт и информационная технология для интеграции распределенных приложений на основе взаимодействия интерфейсов объектов. В большей степени соответствует архитектуре клиент – сервер.

*T<sub>2</sub>. Компонентная (CORBA, CCM).* Дополняет объектную технологию стандартным каркасом приложения и набором стандартизированных интерфейсов (портов), при помощи которых приложения взаимодействуют в распределенной среде.

*T<sub>3</sub>. Технология веб-сервисов (SOAP, WSDL).* Определяет межпрограммное взаимодействие на основе веб-стандартов при помощи веб-сервисов (программных систем, разработанных для поддержки интероперабельного взаимодействия через сеть). Соответствует архитектуре СОА

и является в настоящее время наиболее используемой технологией. Существует развернутый каталог стандартов безопасности веб-сервисов [9], который также включает стандарт интеграции грид-вычислений (OGSI) [10].

*T<sub>4</sub>. Многоагентная технология (KQML, ACL).* Предполагает проектирование самоорганизующейся интеллектуальной системы на основе независимых самообучающихся программных агентов. Агенты общаются между собой при помощи специальных языков (протоколов). В силу слабой защищенности редко применяется для построения РКС.

*T<sub>5</sub>. Технология P2P.* Модель взаимодействия равноправных узлов сети напрямую (без промежуточных серверов и экранов). В таких сетях отсутствуют какие-либо стандарты взаимодействия распределенных приложений; следовательно, данная технология может передавать данные согласно любому принятому протоколу. Соответствует архитектуре одноранговых сетей.

*T<sub>6</sub>. Облачная технология.* Является развитием технологии веб-сервисов. Под этим термином понимается информационно-технологическая концепция (а не технология), гарантирующая наличие общего и повсеместного доступа по требованию к общему пулу вычислительных ресурсов. Согласно стандарту NIST основными свойствами такого пула являются эластичность, учет потребления, объединение ресурсов, самообслуживание по требованию, универсальный доступ по сети. Таким образом, в облачных технологиях используется целый комплекс технологий виртуализации, баланса нагрузки, интеграции ресурсов и приложений. В настоящее время не существует строгих стандартов облачных технологий (в отличие от веб-сервисов) [11], а сам этот термин начал применяться лишь в 2007–2008 гг.

Анализ указанных архитектур и технологий позволяет сделать вывод, что во многом эти компоненты ИС являются взаимосвязанными, поэтому переменная *T* в архитектурно-управленческой модели РКС важна лишь с точки зрения построения систем защиты, а не классификации РКС. Например, если для технологии веб-сервисов и соответствующих ей типов архитектур (клиент – сервер, СОА, грид) существует целая серия стандартов безопасности и сводов практик [9, 10, 12], то в области облачных систем и сервисов наблюдается значительный пробел, который требует своего решения [11, 8]. По этой причине в предложенной модели технологии для различного уровня взаимодействия сетевых сервисов, например CORBA и P2P, объединены в один компонент *T*.

Наконец, третий параметр – управление системами:

*M<sub>1</sub>. Централизованное управление.* Подразумевает наличие в системе одного сервера с ограниченным доступом. Для большинства современных РКС такой вид управления является труднодостижимым идеалом.

*M<sub>2</sub>. Частично централизованное управление.* Означает доступ ограниченного числа лиц и компьютерных систем. Часто встречается в корпоративных системах.

*M<sub>3</sub>. Децентрализованное управление.* Характерно для наиболее открытых типов систем (P2P или общее облако). Безопасность информации в таких системах можно гарантировать лишь с помощью достаточно надежных криптографических методов защиты данных.

С учетом приведенных выше рассуждений любую РКС можно записать как функцию трех параметров:

$$DS = (A_i, T_j, M_k), \quad i = 1, \dots, 5, \quad j = 1, \dots, 6, \quad k = 1, 2, 3.$$

Подставляя в модель различные значения параметров, можно описывать различные конфигурации. Таким образом, концептуальные модели можно использовать для описания различных классов ИС с целью последующего их анализа.

Видно, что наиболее важной частью модели является ее архитектурно-технологическая составляющая. В соответствии с этой составляющей можно выделить пять базовых классов РКС:  $K_1(A_3)$ ,  $K_2(A_1)$ ,  $K_3(A_2)$ ,  $K_4(A_4)$ ,  $K_5(A_5)$ , а с учетом управляющей составляющей – восемь подклассов. Такая классификация распределенных систем позволит определить возможные механизмы, которые в дальнейшем можно использовать для построения системы их защиты. Опишем все указанные классы:

*Класс  $K_1$*  – архитектура одноранговых сетей и технология P2P, в которых возможен лишь один уровень централизации управления:

$M_3$  – полностью децентрализованные системы. В силу отсутствия единого центра управления в таких системах практически невозможно обеспечить приемлемый уровень защищенности.

*Класс  $K_2$*  – многочисленный класс клиент-серверных систем (им соответствует объектная, компонентная технология или технология веб-сервисов). В зависимости от уровня централизации управления можно выделить следующие его подклассы:

$M_1$  – полностью централизованные;

$M_2$  – частично централизованные (при наличии нескольких серверов).

В указанных подклассах необходимо обеспечить двухуровневую систему защиты – для серверов и клиентов. В силу высокого уровня централизации можно сравнительно легко обеспечить защиту таких систем. Основной проблемой остается высокая вероятность сбоев (низкая отказоустойчивость в силу отсутствия распределенной обработки данных) и отсутствие возможностей для масштабирования и расширения.

*Класс  $K_3$*  – сервис-ориентированная архитектура, которой соответствует технология веб-сервисов. В зависимости от уровня централизации управления можно выделить следующий подкласс:

$M_2$  – частично централизованные системы.

В таких системах решение проблемы защиты информации эквивалентно обеспечению защиты сервисов (функций системы). Существуют технологические стандарты безопасности веб-сервисов и стандарты, определяющие механизмы защиты, а также общие принципы построения таких систем защиты. Основной проблемой в данном случае является разрозненность стандартов и повсеместное использование СОА, что требует в ситуациях, не описанных в стандартах, своевременного принятия технологичных и других решений.

*Класс  $K_4$*  – архитектура и технология грид-вычислений. Возможны следующие варианты управления:

$M_2$  – частичное;

$M_3$  – децентрализованное.

Этот класс систем характеризуется разрозненностью стандартов и отсутствием эффективного технологичного решения проблемы безопасности [10, 12].

*Класс  $K_5$*  – облачные системы и технологии. По уровню централизации управления их можно разбить на следующие подклассы:

$M_1$  – частично централизованные;

$M_2$  – полностью децентрализованные.

Основной проблемой в данном случае является отсутствие стандартов и критериев оценки защиты информации [11].

Таким образом, использование архитектурно-управленческих моделей позволяет провести классификацию РКС и выделить подклассы, в рамках которых достижимы различные уровни защиты информации. Классификация систем позволяет также согласовать механизмы защиты с критериями (методиками) решения и определить базис для построения комплексной системы защиты.

## 2. Выбор инфраструктуры безопасности

В основе большинства современных СЗИ лежат механизмы управления ключами, которые определяют базовую совокупность сервисов защиты информации [3, 4, 13], а все дополнительные механизмы и службы безопасности включаются в СЗИ уже в качестве ее надстройки. Ядром механизма управления ключами, как правило, является криптографический протокол проверки подлинности (ППП), который определяет концепцию доверия [13, 14]. Таким образом, можно считать, что механизм управления ключами (МУК) и ППП в совокупности определяют своеобразный базис СЗИ, т. е. своеобразную инфраструктуру ее безопасности.

Под инфраструктурой безопасности (ИБ) в дальнейшем будем понимать программно-аппаратную технологию, которая определяет механизмы аутентификации пользователей системы на основе ППП, МУК и механизмы согласования ключей между пользователями и сервисами системы.

Заметим, что понятие «инфраструктура безопасности» в различных литературных источниках именуется по-разному. Например, в монографии [14] оно определяется как «инфраструктура открытых ключей» (один из видов ИБ), а в монографии [3] серверы ключей и их функции неявно также удовлетворяют необходимым требованиям, а значит, являются прототипами ИБ. Можно также отметить, что отсутствует и какая-либо классификация технологий ИБ.

Вместе с тем анализ подходов к решению проблемы аутентификации и управления ключами позволяет сделать вывод, что в качестве базового множества ИБ можно принять следующие типы инфраструктур:

КИ<sub>1</sub> – парольная инфраструктура;

КИ<sub>2</sub> – инфраструктура протокола Kerberos;

КИ<sub>3</sub> – инфраструктура открытых ключей;

КИ<sub>4</sub> – инфраструктура открытых ключей (корпоративная);

КИ<sub>5</sub> – собственная инфраструктура согласования ключей и функции аутентификации.

Каждая из перечисленных инфраструктур обладает как определенными достоинствами, так и недостатками и подходит только для конкретного типа архитектуры и уровня централизации РКС. Наиболее передовыми и надежными считаются технологии, связанные с инфраструктурой открытых ключей [14]. Однако их реализация в системах типа К<sub>2</sub> (клиент – сервер), как правило, является трудоемкой и неэффективной. В то же время инфраструктуры типа КИ<sub>3</sub>, КИ<sub>4</sub> или КИ<sub>5</sub> более пригодны для систем с удаленным доступом, т. е. для облачных технологий [11] (архитектур К<sub>3</sub>, К<sub>4</sub>, К<sub>5</sub>) или для грид-технологий [12] и веб-сервисов [9]. В корпоративных сетях, основанных, как правило, на архитектуре К<sub>2</sub> или К<sub>3</sub>, менее трудоемкой и независимой от времени доступа является реализация протокола и технологии управления ключами Kerberos (КИ<sub>2</sub>) или (при отсутствии у пользователя конфиденциальных данных (КИ<sub>1</sub>)) парольная инфраструктура:

	К <sub>1</sub>	К <sub>2</sub>	К <sub>3</sub>	К <sub>4</sub>	К <sub>5</sub>
КИ <sub>1</sub>	+	+	–	–	–
КИ <sub>2</sub>	–	+	+	–	–
КИ <sub>3</sub>	–	–	+	+	+
КИ <sub>4</sub>	+	–	–	+	+
КИ <sub>5</sub>	+	+	+	+	+

Выбор инфраструктуры безопасности зависит также от степени централизации РКС. В централизованных, как правило, более защищенных системах достаточно присутствия единого сервера ключей Kerberos (КИ<sub>2</sub>) или, в исключительных случаях, парольной аутентификации КИ<sub>1</sub>. Для более распределенных систем следует использовать инфраструктуру КИ<sub>3</sub> стороннего производителя (при малых ресурсах) или собственную внутреннюю инфраструктуру открытых ключей КИ<sub>4</sub> (при большем количестве ресурсов). В случае большой степени распределенности данных и уровней доступа к системе возможно гибридное решение типа КИ<sub>5</sub>:

	КИ <sub>1</sub>	КИ <sub>2</sub>	КИ <sub>3</sub>	КИ <sub>4</sub>	КИ <sub>5</sub>
М <sub>1</sub>	+	+			+
М <sub>2</sub>	–	±	+	+	+
М <sub>3</sub>	–	–	±	+	+

Выбор конкретной реализации зависит как от объемов располагаемых ресурсов, так и от степени конфиденциальности защищаемой информации. В нашей классификации трудоемкость реализации возрастает от КИ<sub>1</sub> к КИ<sub>5</sub>. Защищенность сервера ключей, наоборот, уменьшается от КИ<sub>5</sub> к КИ<sub>1</sub>.

Очевидно, что серверы ключей [3] независимо от реализации лишь частично решают проблему защиты информации в РКС. Например, они не решают важную задачу авторизации и организации списков доступа [3, 14]. Кроме того, в любой системе должен быть определен механизм аудита или протоколирования процессов, организации дополняющего (скрывающего) трафика, маршрутизации и т. д. [4, 5]. Для решения этих задач, так же как и для уточнения базовой инфраструктуры, очевидно, необходимо провести уточнение архитектурно-управленческой модели на более низком компонентном уровне.

Таким образом, предложенная выше модель ИС позволяет определить только принципиальную инфраструктуру механизмов защиты и общий подход к оценке защищенности. Очевидно, что такую модель целесообразно использовать на этапе инициализации комплексной системы защиты. В большинстве случаев ее параметров недостаточно для определения конкретной конфигурации основных и дополнительных механизмов защиты, необходимых для отдельных элементов ИС (узлов, ресурсов, процессов сети и ее инфраструктуры).

### 3. Компонентная модель РКС

Формальный способ описания распределенных систем в виде концептуальных моделей позволил определить специализированную методологию построения системы их защиты. Для решения общей задачи защиты информации, как правило, применяется система стандартов ISO/IEC 15408 (В РБ СТБ 34.101.1-3) [15-17]. Однако для того чтобы согласовать конкретную РКС с набором ее средств защиты, необходимо перейти от общего описания классов к их детальному наполнению. На важность компонентного подхода в решении задачи безопасности для открытых систем указывает и специализированный стандарт сетевой безопасности под названием «Интерпретация общих критериев защищенных систем для сетевых конфигураций» [18]. Основными компонентами (объектами) распределенной системы, как правило, являются:

- ресурсы ( $R$ );
- узлы ( $N$ );
- процессы ( $Pr$ ).

Таким образом, концептуальную модель системы на компонентном уровне можно представить в виде следующего кортежа:

$$C = (R, N, Pr).$$

К ресурсам относятся программный код или аппаратные средства, используемые в РКС:

- данные, хранящиеся в системе;
- сервисы, т. е. функциональные возможности, которые система предоставляет пользователям;
- узлы сети за исключением клиентов, т. е. потребителей информации;
- коммуникационные каналы сети.

К узлам системы относятся все подключенные к ней устройства. При этом в зависимости от типа архитектур возможны несколько типов таких устройств:

- клиенты, т. е. потребители информации и сервисов;
- серверы, т. е. устройства, обеспечивающие функциональность сети;
- устройства хранения данных;
- пиры – узлы, которые могут иметь функциональность на уровне клиента или сервера.

Процессами в системе являются любые исполняемые на узлах сети и передаваемые по ее коммуникационным каналам программные коды:

- информационные потоки сервисов;
- пользовательские процессы (сервисы);
- внутренние процессы управления системой;
- защитные процессы.

В качестве основной единицы процесса моделирования будем рассматривать узлы сети.

Пусть в сети имеется  $k$  узлов:

$$N = (N_1, \dots, N_k).$$

Предположим, что в узлах сети хранятся некоторые наборы данных, т. е. каждому узлу  $N_i$  соответствует набор данных  $(D_1, \dots, D_{i(i)})$ :

$$N_i \rightarrow D_i = (D_1, \dots, D_{i(i)}).$$

Переменные  $D_j$  означают метки безопасности данных узла. Метки безопасности определяют уровни секретности, поддерживаемые системой, и образуют упорядоченные множества, например: «совершенно секретно», «секретно», «несекретно» и «для общего доступа». Отметим, что в общем случае метки безопасности должны быть приведены в соответствие с СТБ 34.101.30 [6]. В различных системах категории доступа и наборы меток безопасности могут отличаться друг от друга. Таким образом, набору данных в системе в общем случае соответствует матрица  $k(t)$ , где  $t$  – общее число меток безопасности.

Очевидно, матрица каналов для  $k$  узлов связи содержит  $k^2$  значений. Каждому каналу, соединяющему узлы  $N_i$  и  $N_j$ , присваивается метка его безопасности:

$$Ch = (Ch_{11}, \dots, Ch_{k,k}).$$

При наличии между узлами нескольких каналов их также можно задать наборами меток.

Для определения процессов и сервисов системы введем функцию  $Pr$  (*Send*, *Receive*), где компонента *Send* определяет вектор, задающий подмножество узлов  $(N_{i1}, \dots, N_{is})$ , которые передают матрицу данных  $(D_{i1}, \dots, D_{is})$ , а компонента *Receive* – соответственно вектор и матрицу данных узлов-получателей. Таким образом, в общем случае процесс представляется в виде двух подматриц набора данных системы.

С учетом приведенных рассуждений каждую систему на данном компонентном уровне можно задать в виде тройки

$$SDS = (D, Ch, Pr),$$

где  $D$  определяет матрицу меток безопасности данных узлов сети,  $Ch$  – матрицу меток безопасности каналов сети, а  $Pr$  – совокупность пар подматриц матрицы  $D$ .

Видно, что описанная выше модель позволяет определить необходимый уровень безопасности каждого компонента системы, поскольку требуемый уровень (профиль) безопасности определяется новым защищенным состоянием системы:

$$SDS_{pr} = (D_{pr}, Ch_{pr}, Pr_{pr}),$$

где  $D_{pr}$  – набор защищенных данных системы, т. е. измененная матрица меток безопасности с учетом функциональности добавленных механизмов. Пусть при анализе системы выяснилось, что в некотором узле в открытом доступе хранятся данные, которые нужно сделать конфиденциальными. После реализации механизма шифрования защищенность этих данных повысится, а значит, соответствующее значение в матрице меток поменяется;

$Ch_{pr}$  – набор защищенных каналов. После добавления функциональности механизмов безопасности в системе могут появиться новые, более защищенные каналы связи, которые могут быть организованы при наличии базовой инфраструктуры при помощи безопасного обмена ключами симметричного или асимметричного шифрования;

$Pr_{pr}$  – набор процессов защищенной системы, который также включает в себя процессы, относящиеся к механизмам безопасности.

Результатом решения задачи защиты информации (ЗЗИ) является переход РКС из текущего состояния  $SDS$  в состояние  $SDS_{pr}$ . Видно, что модель РКС на компонентном уровне позволяет оценить текущее состояние системы  $SDS$  и выбрать необходимые механизмы безопасности с целью реализации целостной ЗЗИ, обеспечивающей состояние системы  $SDS_{pr}$ .

Отметим, что принятие решения о выборе механизмов безопасности является, по сути, оптимизационной задачей, поскольку на выбор механизмов влияет не только вероятность осуществления угроз, но и стоимость затрат на реализацию механизмов. Таким образом, предложенный подход к решению задачи обеспечения защищенного состояния РКС тесно связан с традиционной методикой оценки рисков безопасности.

Таким образом, иерархия моделей *DS* и *SDS* (на архитектурно-управленческом и компонентном уровнях) дает возможность определять объекты защиты, их основные характеристики и требуемый уровень защиты, что позволит ставить и решать ЗЗИ в распределенной системе. В первую очередь для решения данной задачи необходимо определить инструменты или средства защиты РКС, которые базируются на выбранной инфраструктуре безопасности в соответствии с классом РКС.

#### 4. Механизмы и функции обеспечения безопасности

Для того чтобы построить систему защиты, необходимо определить не только объект защиты, но и средства защиты [4]. Объектом защиты является модель распределенной системы *SDS* или кортеж компонентов системы в совокупности с характеристикой класса системы *DS*. Механизмы безопасности (*Meh*) по признаку принадлежности к сервисам безопасности (семейству взаимосвязанных механизмов для предоставления определенной защитной функции) делятся на следующие группы:

- аутентификация;
- управление доступом;
- конфиденциальность данных;
- целостность данных;
- предотвращение отказа от авторства.

Помимо сервисов имеются непосредственные механизмы безопасности, которые могут относиться к сервисам или поставляться отдельно:

- шифрование;
- цифровая подпись;
- механизмы разграничения доступа;
- механизмы контроля целостности;
- протоколы взаимной аутентификации;
- подстановка трафика;
- управление маршрутизацией;
- механизм заверения;
- доверенная функциональность;
- метки безопасности;
- обнаружение событий;
- аудит безопасности;
- восстановление безопасного состояния.

Для того чтобы выбрать набор механизмов или средств безопасности для конкретной распределенной системы из доступного множества, необходимо сформировать базовое множество алгоритмических (*Alg*) и технических (*Prot*) средств [3, 4, 11]. Каждый из этих механизмов, по сути, представляет собой совокупность процессов, в которых принимают участие компоненты (узлы, данные, процессы) системы, т. е. которым обеспечивается определенное свойство системы на компонентном уровне. Например, аутентификация – это поддержка процессов проверки подлинности пользователей системы. Таким образом, в наиболее общем виде механизм *M* можно описать как вектор (кортеж)

$$Meh = (T, Pr_1, Pr_2, \dots, Pr_n, Scene),$$

где *T* – применяемая технология (*Alg* или *Prot*);  $\{Pr_1, Pr_2, \dots, Pr_n\}$  – совокупность процессов механизма; *Scene* – совокупность участников функционирования механизма, ресурсов ( $D_{scene}$ ), каналов ( $Ch_{scene}$ ) и процессов ( $Pr_{scene}$ ), т. е. часть системы на компонентном уровне, имеющая отношение к реализации или протокола:

$$Scene = (D_{scene}, Ch_{scene}, Pr_{scene}).$$

Результатом действия механизма является изменение состояния сцены, т. е.  $Meh(T, Pr_1, \dots, Pr_n, Scene) = Scene_{pr}$ , где  $Scene_{pr}$  – защищенное состояние подмножества компонентов системы. Видно, что в результате действия последовательности механизмов  $Meh_1, \dots, Meh_k$  вся система

перейдет в состояние  $SDS_{pr}$ . Следует отметить, что ключевыми в сетевых конфигурациях являются криптографические механизмы ( $T = Alg$ ), поскольку именно они гарантируют доступ к данным независимо от операционной системы и устройства.

Выбор механизмов безопасности зачастую представляет собой задачу принятия решения. Для решения подобных задач, как правило, используются экспертные оценки и методы.

## 5. Методика построения СЗИ

В основе описанного выше концептуального подхода лежит выбор средств защиты в зависимости от класса и компонентного состава РКС. Такая методология обобщает известные ранее подходы к защите информации открытых систем, основанные прежде всего на выборе конкретных механизмов, ориентированных на решение некоторых подзадач ЗЗИ, т. е. покрывающих лишь часть необходимой функциональности комплексной СЗИ [13, 14]. Это затрудняло поиск наиболее эффективного решения для данного класса систем. В соответствии с обобщенной формулировкой ЗЗИ (см. разд. 2) предлагается следующая поэтапная методика проектирования СЗИ:

1. Идентификация класса системы, т. е. определение параметра DS.
2. Выбор класса инфраструктуры безопасности и базовой комплектации сервисов безопасности.
3. Идентификация компонентов системы, т. е. определение параметра SDS.
4. Уточнение требований к уровню защиты ( $SDS_{pr}$ ).
5. Инициализация механизмов защиты  $M_i$  в зависимости от выбранных технологий и в соответствии с достижимым уровнем защиты.
6. Настройка параметров механизмов на конкретную систему (решение оптимизационной задачи).

Необходимым условием проектирования является аппаратная и системная независимость такой СЗИ. Следовательно, наиболее предпочтительными средствами и методами защиты служат криптографические механизмы [13], надежность которых не зависит от каналов связи, узлов сети и доступа конечных пользователей.

Прикладные аспекты построения комплексной системы защиты ранее не являлись актуальной ЗЗИ, поскольку не все компоненты РКС находились в открытом доступе, а систем, в полной мере удовлетворяющих характеристикам распределенности, практически не существовало. В нынешних условиях необходимость в создании аппаратно-, платформи- и каналонезависимых средств защиты возрастает в силу значительного роста популярности и использования облачных систем и сервисов, которые, по сути, являются открытыми и децентрализованными системами [8, 11].

Применение методики идентификации системы с последующей оптимизационной настройкой параметров позволит эффективно и точно решать проблему безопасности открытых систем в короткие сроки.

Рассмотрим решение проблемы защиты информации на примере ИС с распределенной облачной инфраструктурой. Пусть известен исходный класс системы (облачный), тип управления (частично централизованный, корпоративный) и основные цели системы (хранение данных, управление доступом). В соответствии с описанным выше подходом можно достаточно точно определить настройки системы безопасности.

Первый уровень иерархии – *архитектурный*. Он включает следующие этапы:

1. Выбор типа аутентификации. Для рассматриваемой системы хранения данных необходима двухфакторная аутентификация. Так как в системе имеются данные, доступ к которым нужно ограничивать, аутентификацию предлагается проводить с помощью пароля и доступа к почтовому ящику (мобильному устройству).

2. Выбор инфраструктуры безопасности. Для систем с облачной инфраструктурой необходимо использовать уровень защиты не ниже шифрования с открытым ключом. Поскольку требуется шифровать большие объемы данных, принято решение использовать гибридную систему шифрования: для защиты пользовательских сессионных ключей – инфраструктуру PKI, а для шифрования самих файлов – AES128.

Второй уровень иерархии – *компонентный*. Настройка дополнительных механизмов в разных системах осуществляется по-разному:

1. В облачных сервисах авторизация должна быть независимой от платформы и типа устройства. Кроме того, на выбор типа авторизации влияют дополнительные факторы, такие как скорость шифрования, тип организации пользовательских групп доступа, стоимость реализации и т. д. С учетом этого был выбран оптимальный для таких сервисов тип авторизации: шифрование с атрибутами, модифицированное с учетом особенностей системы.

2. На сервере организуется аудит всех событий, происходящих в системе.

Описанная концепция безопасности реализована в рамках выполнения реального прикладного проекта. Базовая конфигурация механизмов была выбрана в соответствии с двухуровневой методологией; система безопасности состояла из четырех компонентов – клиента, сервера и хранилища файлов и ключей. Сервер в данном случае призван поддерживать функциональность ИС, а клиент – защищать данные на устройствах пользователей, при этом клиент с хранилищем файлов и ключей напрямую не взаимодействует.

### Заключение

В работе описан один из возможных подходов к построению СЗИ в РКС. Рассмотрены основные архитектуры распределенных систем, соответствующие механизмы безопасности и принципы построения СЗИ. Построена иерархия формальных (концептуальных) моделей, которая позволяет описывать различные классы и конфигурации ИС. На первом (архитектурно-управленческом) уровне иерархии определяется база СЗИ, т. е. инфраструктура безопасности, которая включает способы аутентификации и установления безопасного соединения. Более тонкие механизмы, к которым относятся авторизация, настройка маршрутов и аудит данных, определяются на втором (компонентном) уровне.

В отличие от существующих подходов, которые опираются на выбор совокупности независимых механизмов, предлагаемая методология построения СЗИ носит специализированный характер. Также описана целостная концепция безопасности систем с облачной инфраструктурой, указана пошаговая стратегия выбора и конфигурирование механизмов безопасности.

Таким образом, в работе предложена методология, которая позволяет упростить и автоматизировать процесс проектирования СЗИ в соответствии с целями и особенностями РКС. Результаты работы обсуждались и получили признание на Международной научной конференции консорциума IEEE [19].

### Список литературы

1. Таненбаум, Э. Распределенные системы. Принципы и парадигмы / Э. Таненбаум, М. ван Стеен. – СПб. : Питер, 2003. – 877 с.
2. Косяков, М.С. Введение в распределенные вычисления / М.С. Косяков. – СПб. : НИУ ИТМО, 2014. – 155 с.
3. Фергюсон, Н. Практическая криптография / Н. Фергюсон, Б. Шнайер. – М. : Диалектика, 2004. – 432 с.
4. О некоторых вопросах технической и криптографической защиты информации : Приказ Оперативно-аналитического центра при Президенте Республики Беларусь, 30 авг. 2013 г. № 62 // Нац. реестр правовых актов Респ. Беларусь. – 2015. – № 3. – 22 с.
5. Галатенко, В.А. Стандарты в области безопасности распределенных систем / В.А. Галатенко // Jet Info. – 1999. – № 5. – С. 98–99.
6. Информационные технологии. Методы и средства безопасности. Объекты информатизации. Классификация : СТБ 34.101.30–2007. – Введ. 01.04.08. – Минск : Межгос. совет по стандартизации, метрологии и сертификации : Белорус. гос. ин-т стандартизации и сертификации, 2007. – 7 с.
7. Радченко, Г.И. Распределенные вычислительные системы / Г.И. Радченко. – Челябинск : Фотохудожник, 2012. – 184 с.
8. Galibus, T. Cloud Storage Security / T. Galibus, H. Vissia // Proc. of NSCE'2014, 24–25 December 2014. – Hong Cong : CRC Press, 2015. – P. 123–127.

9. OASIS Web Services security Technical Committee and standardization group [Electronic resource]. – 2015. – Mode of access : [https://www.oasis-open.org/committees/tc\\_home.php?wg\\_abbrev=wss](https://www.oasis-open.org/committees/tc_home.php?wg_abbrev=wss). – Date of access : 19.10.2015.
10. Open Grid Services Infrastructure (OGSI) Specification Version 1.0 [Electronic resource]. – 2003. – Mode of access : <http://xml.coverpages.org/OGSI-SpecificationV110.pdf>. – Date of access : 21.10.2015.
11. Cloud Security Alliance, organization promoting the use of best practices for providing security assurance within Cloud Computing [Electronic resource]. – 2015. – Mode of access : <https://cloudsecurityalliance.org>. – Date of access : 11.10.2015.
12. A Security Architecture for Computational Grids / I. Foster [et al.] // Proc. 5th ACM Conf. on Computer and Communications Security. – N.Y., 1998. – P. 83–92.
13. Шнайер, Б. Прикладная криптография. Протоколы, алгоритмы, исходные тексты на языке Си / Б. Шнайер. – М. : Триумф, 2002. – 816 с.
14. Полянская, О.Ю. Инфраструктуры открытых ключей / О.Ю. Полянская, В.С. Горбатов. – М. : Бино, 2007. – 216 с.
15. Информационные технологии и безопасность. Критерии оценки безопасности информационных технологий. Часть 1. Введение и общая модель : СТБ 34.101.1–2014 (ISO/IEC 15408–1:2009). – Введ. 01.09.14. – Минск : Межгос. совет по стандартизации, метрологии и сертификации : Белорус. гос. ин-т стандартизации и сертификации, 2014. – 53 с.
16. Информационные технологии и безопасность. Критерии оценки безопасности информационных технологий. Часть 2. Функциональные требования безопасности : СТБ 34.101.2–2014 (ISO/IEC 15408–2:2009). – Введ. 01.09.14. – Минск : Межгос. совет по стандартизации, метрологии и сертификации : Белорус. гос. ин-т стандартизации и сертификации, 2014. – 178 с.
17. Информационные технологии и безопасность. Критерии оценки безопасности информационных технологий. Часть 3. Гарантийные требования безопасности : СТБ 34.101.3–2014 (ISO/IEC 15408-3:2009). – Введ. 01.09.14. – Минск : Межгос. совет по стандартизации, метрологии и сертификации : Белорус. гос. ин-т стандартизации и сертификации, 2014. – 131 с.
18. Trusted Network Interpretation NIST security standard [Electronic resource]. – 2015. – Mode of access : <http://csrc.nist.gov/publications/secpubs/rainbow/tg005.txt>. – Date of access : 17.10.2015.
19. Krasnoproshin, V. Conceptual Distributed System Models and Organization of Security Mechanisms / V. Krasnoproshin, T. Galibus // Proc. of 8th Intern. IEEE Conf. on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Application (AIDAACS'2015), 24–26 September. – Warsaw, Poland, 2015. – P. 432–43.

Поступила 13.11.2015

*Белорусский государственный университет,  
Минск, пр. Независимости, 4  
e-mail: tan2tan@gmail.com  
krasnoproshin@bsu.by*

**T. Galibus, V. Krasnoproshin**

## **CONCEPTUAL MODELLING AND ORGANIZATION OF SECURITY MECHANISMS IN DISTRIBUTED SYSTEMS**

We analyze the existing DS from the point of security and construct a two-level hierarchy of models. Such approach allows us to separate the abstraction (architecture) level and the concrete (component) level of ISS. The core set of methods, i. e. authentication and key exchange protocols, corresponds to the abstraction level and is defined as security infrastructure (SI). The final security parameters optimization and additional mechanisms such as authorization, routing and data auditing of the protection mechanisms are configured on the component level of the DS. In addition, we outline the systematic step-by-step ISS configuration method.

УДК 004.089

А.И. Трубей

## АНАЛИЗ ОСОБЕННОСТЕЙ ОБЕСПЕЧЕНИЯ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ ГРИД-СИСТЕМ

*Предлагается краткий обзор особенностей обеспечения информационной безопасности грид-систем, рассматривается методика оценки рисков, угроз и уязвимостей с использованием общедоступных баз данных, экспертных систем и статистических методов.*

### Введение

В настоящее время развитие вычислительных технологий достигло своего пика, когда простое увеличение мощности процессоров, объема хранилищ данных, пропускной способности сети передач практически исчерпало свои ресурсы и не удовлетворяет растущим потребностям к вычислительным ресурсам как в науке, образовании, так и в решении сложных прикладных задач. Скорость сетевого взаимодействия становится все более сопоставимой со скоростью процессоров, и, соответственно, вопрос о локализации вычислительных мощностей постепенно теряет свою остроту. В связи с этим наблюдается тенденция сдвига в сторону распределенных вычислительных систем, которые обеспечивают доступ пользователей к географически распределенным вычислительным ресурсам, объединенным в единый мощный пул ресурсов.

Распределенные вычислительные системы уже давно применяются для решения ресурсоемких задач, но в последние годы стали изменяться масштаб и сам характер их использования. Дальнейшее развитие в этой области в будущем может позволить крупным промышленным предприятиям, научно-исследовательским организациям, а также рядовым пользователям получать доступ к гибкой, географически распределенной и масштабируемой инфраструктуре для обработки и хранения данных. Существует большое количество разнообразных видов и классификаций распределенных систем, таких как кластерные системы, грид-системы, облачные системы, пиринговые сети (peer-to-peer, P2P) и т. д.

Очевидно, что при разработке подобных сложных систем одной из важнейших задач является реализация механизмов обеспечения безопасности, связанных с необходимостью противодействия компьютерным атакам на вычислительные и информационные ресурсы грид-систем. Особые свойства грид-систем (гетерогенность, перераспределение ресурсов, высокая динамика состояний, децентрализованность управления и т. д.) создают благоприятные условия для осуществления на них таких негативных воздействий, как атаки отказа в обслуживании, распространение вредоносного программного обеспечения, несанкционированный доступ к ресурсам.

В статье детализированы меры, которые необходимо принимать для обеспечения безопасного управления в грид-системе, предложена методика оценки рисков, угроз и уязвимостей с использованием базы общеизвестных уязвимостей, имеющих строгую характеристику по описательным критериям (common vulnerabilities and exposures, CVE) и банка данных Федеральной службы по техническому и экспортному контролю России (ФСТЭК России). Кроме того, приведен метод оценки инцидентов информационной безопасности (ИБ) на основе использования критерия  $\chi^2$ .

### 1. Обеспечение безопасного управления в грид-системе

Безопасное управление требует контроля доступа к услугам через надежные протоколы безопасности и в соответствии с принятой политикой безопасности. Например, получение прикладных программ и запуск их в грид-среде может потребовать аутентификации и идентификации. Совместное использование ресурсов пользователями требует наличия какого-нибудь вида механизма изоляции. Кроме того, необходимы стандартные механизмы безопасности, которые могут быть направлены на защиту грид-систем путем поддержки безопасного совместного использования ресурсов между административными доменами [1].

Эффективность принимаемых мер защиты информации зависит от качества определения угроз безопасности информации конкретной грид-системы в конкретных условиях ее функционирования. Они, как правило, включают следующие меры защиты:

**1. Аутентификацию и идентификацию.** Для идентификации индивидуальных пользователей и установления запрашиваемых сервисов необходимы соответствующие механизмы аутентификации. Грид-система должна удовлетворять требованиям политики безопасности каждого домена, а также, возможно, определять пользовательские политики безопасности. Аутентификация и идентификация должны включать различные модели контроля доступа и их реализации [1]. Особенности аутентификации и идентификации в грид-системе являются единый вход и делегирование прав.

*Единый вход.* Пользователь должен аутентифицироваться только один раз в начале сеанса работы, получая доступ ко всем разрешенным ресурсам базового уровня архитектуры грид-системы. Технология единого входа позволяет избежать повторной процедуры аутентификации во вторичных доменах. Применение данного подхода:

- сокращает время, затрачиваемое пользователем на аутентификацию в каждом домене;
- повышает безопасность за счет сокращения аутентификационной информации, которую необходимо запоминать пользователю;
- сокращает время, затрачиваемое системным администратором на управление учетными записями;
- сокращает нагрузку на серверы доступа за счет уменьшения количества процедур аутентификации.

*Делегирование прав (доверия к аутентификатору).* Важным условием для эффективной работы распределенных систем является возможность делегирования прав пользователя грид-сервисам. Практически любой запрос пользователя проходит через несколько сервисов. При отсутствии механизма делегирования пользователю было бы необходимо аутентифицироваться на каждом сервисе в цепочке сервисов, обрабатывающих данный запрос, т. е. пользователь после отправления задания должен постоянно отвечать на запросы о своей аутентификации от каждого сервиса в обрабатывающей цепочке. Это усложняет работу в грид-системе – во всяком случае при запуске большого набора заданий. Делегирование прав позволяет этого избежать. Пользователь должен иметь возможность запуска программ от своего имени. Программы получают доступ ко всем ресурсам, к которым имеет доступ пользователь, могут делегировать часть своих прав другим программам. Риск злоупотребления делегированными правами должен быть сведен к минимуму, например, путем ограничения прав, переданных посредством делегирования на запланированную работу, и ограничения длительности их существования.

В настоящее время существует инфраструктура безопасности грид-систем (grid security infrastructure, GSI), которая обеспечивает безопасную работу в незащищенных сетях общего доступа, предоставляя такие сервисы, как аутентификация, конфиденциальность передачи информации и единый вход в грид-систему. Поскольку вычислительной платформой для решения пользовательских задач в грид-системе являются связанные компьютеры, а не серверы (как это реализовано в кластерных системах и системах облачных вычислений), важной задачей обеспечения защиты данных пользователей грид-систем от атак несанкционированного доступа является изоляция указанных данных от воздействия факторов хостовой среды [2]. Инфраструктура GSI основана на надежных и широко используемых сервисах инфраструктуры криптографии с открытым ключом (public key infrastructure, PKI), которые наиболее подробно описаны в рекомендациях ITU-T X.842 [3].

Основными механизмами аутентификация в PKI являются:

- установление доверия на базе определенной модели доверия;
- система именования субъектов, обеспечивающая уникальность имени в рамках данной системы;
- связь имени субъекта и пары ключей (открытый и закрытый) с подтверждением этой связи средствами удостоверяющего центра (УЦ), которому доверяет субъект, проверяющий правильность связи.

Для обеспечения высоких уровней доверия к результатам аутентификации при доступе сотрудников государственных организаций к информационным системам (в том числе и к гирд-ресурсам) все чаще используются так называемые смарт-карты с неизвлекаемым секретным ключом [4]. Они способны безопасно генерировать внутри чипа ключевой материал. Устройства генерируют две ключевые пары. Первая пара используется для организации доступа к сети и информационным ресурсам. На открытый ключ первой пары УЦ выпускается сертификат, который применяется для управления доступом. Вторая пара ключей используется для электронно-цифровой подписи (ЭЦП): открытый ключ пересылается в УЦ для формирования сертификата ключа проверки подписи на основе тщательной идентификации владельца, все этапы которой протоколируются.

Неизвлекаемость закрытых ключей гарантируется требованиями к чипам смарт-карт, закрепленными Федеральными стандартами обработки информации (federal information processing standards, FIPS), в частности стандартом FIPS-140 и уровнем доверия не ниже EAL4+ (уровнем гарантии оценки УГО4+ в гармонизированном стандарте СТБ 34.101.3–2014 [5]). Аутентификатор представляет собой аппаратный криптографический модуль, прошедший проверку соответствия FIPS-140-2 уровня 2 и выше. Для аутентификации требуется, чтобы заявитель по безопасному протоколу аутентификации подтвердил принадлежность данного аутентификатора именно ему, а также возможность управления аутентификатором.

Неизвлекаемость секретного ключа означает то, что он не покидает чипа, внутри которого его сгенерировали в составе ключевой пары. Все внешние функции (запрос на сертификат, проверку подписи и т. д.) выполняет открытый ключ. Частным случаем применения таких устройств является решение одной из самых сложных задач – проблемы неотказуемости от подписи с помощью ЭЦП документа или сообщения.

Наиболее популярным способом аутентификации пользователей является аутентификация по паролю. Одним из направлений усовершенствования этого способа может стать применение усиленной аутентификации путем добавления дополнительных признаков, в том числе ритма набора пароля, например, по ритму определенной мелодии. На первоначальном этапе пользователь вводит пароль с заданным ритмом, интервалы между нажатиями клавиш сохраняются в памяти компьютера как эталонный образец и при последующих входах этот образец сравнивается с новыми наборами пароля [6].

Многофакторная, или расширенная, аутентификация основана на совместном использовании нескольких факторов аутентификации (знаний, средств или объектов хранения одной из информационных составляющих легитимной процедуры аутентификации), что значительно повышает безопасность использования информации по меньшей мере со стороны пользователей, подключающихся к информационным системам по защищенным и незащищенным каналам коммуникаций. В настоящее время самым популярным вторым фактором, используемым поставщиками сервиса, является одноразовый пароль (one time password, OTP). Кроме того, в качестве дополнительной опции для определения прав доступа к информационным ресурсам используются биометрические данные человека (отпечатки пальцев, изображение лица и т. д.). Главное преимущество биометрической идентификации заключается в том, что идентифицируется конкретный человек, а не отчужденный носитель (карта, жетон и т. д.) или пароль. Биометрический идентификатор нельзя забыть, украсть или передать [7].

**2. Обмен политиками безопасности.** Потребители и поставщики сервисов должны иметь возможность динамически обмениваться информацией о политиках безопасности, чтобы установить путем переговоров контекст безопасности между ними [1]. Обмен политиками безопасности – предоставление возможности обмена информацией о политиках безопасности вызывающей и вызываемой сторонам для создания безопасной среды обмена информацией. Каждый ресурс может использовать любой из существующих способов решения проблемы безопасности системы. Реализация системы безопасности гирд-систем должна взаимодействовать с этими локальными системами безопасности, не изменяя их. Пользователи должны иметь возможность создавать новые сервисы (ресурсы) динамически, без вмешательства администратора. Эти сервисы должны координироваться и безопасно взаимодействовать с другими сервисами. Доступ к ним должен предоставляться без противоречий с локальной политикой безопасности.

## 2. Предотвращение вторжений, оценка рисков, угроз и уязвимостей с использованием общедоступных баз данных

В соответствии с приказом Оперативно-аналитического центра при Президенте Республики Беларусь от 16.01.2015 № 3 [8] «в целях реализации политики информационной безопасности разрабатываются локальные нормативные правовые акты организации, регламентирующие порядок:

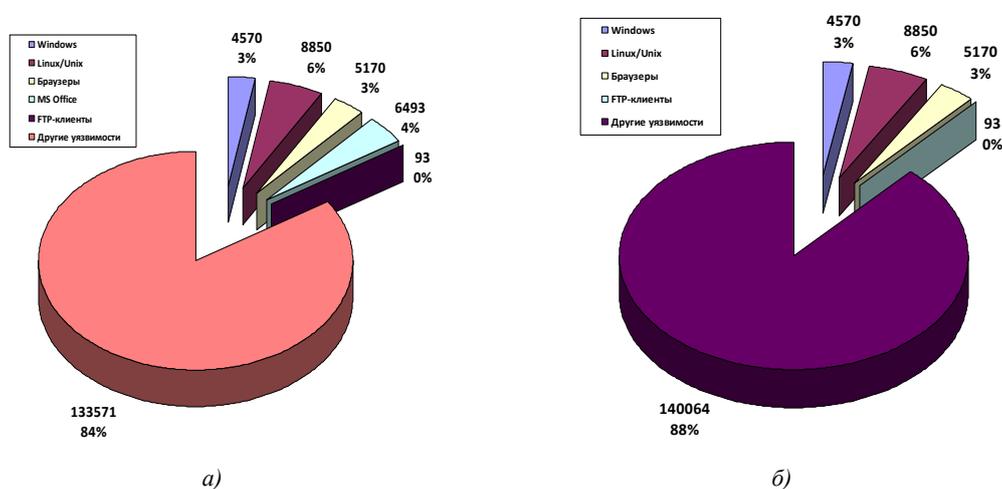
- защиты от вредоносного программного обеспечения;
- выявления угроз, которые могут привести к сбоям, нарушению функционирования информационной системы;
- реагирования на инциденты информационной безопасности...».

Угрозы безопасности информации в грид-системе определяются по результатам оценки возможностей (потенциала, оснащенности и мотивации) внешних и внутренних нарушителей, анализа возможных уязвимостей, способов реализации угроз безопасности информации и последствий от нарушения свойств безопасности информации. В грид-системе должно обеспечиваться обнаружение (предотвращение) угроз, вторжений (компьютерных атак), направленных на преднамеренный несанкционированный доступ к информации, специальных воздействий на информацию (носители информации) в целях ее хищения, уничтожения, искажения и блокирования доступа к ней [9]. Для обнаружения и идентификации злоупотреблений (в том числе действий вирусов) необходима мощная система мониторинга. Для защиты критически важных областей или функций необходимо обеспечить возможность перемещения этих областей для отвода от них атаки [1].

Адекватным ответом на разветвленный и многоуровневый характер процессов реализации угроз в грид-системах, огромное количество пользователей, изменчивость их состава и предоставляемых им услуг является создание так называемых многоагентных систем защиты. В настоящее время создаются системы защиты с централизованным принципом построения, а в многоагентных системах в основном применяется централизованно-децентрализованный принцип построения.

При централизованном принципе все элементы системы защиты замыкаются на консоль администратора безопасности. В многоагентных системах так называемые агенты (программные или программно-аппаратные компоненты) реализуют предопределенные функции безопасности. Они взаимодействуют между собой, обмениваются сообщениями на языке высокого уровня для принятия согласованных решений, адаптируются к реконфигурации сети, изменению трафика и новым видам атак и иным нарушениям безопасности информации, используют алгоритмы обучения, инициализируют деятельность других агентов и решают, таким образом, весь необходимый комплекс задач защиты информации в информационной системе [10].

Для идентификации и анализа уязвимостей и угроз, как правило, используется база общеизвестных уязвимостей CVE [10, 11] (рисунки).



Структура уязвимостей согласно базе CVE: а) в грид-системах; б) в суперкомпьютерных системах

Кроме того, для анализа уязвимостей и угроз, наиболее характерных для государственных информационных систем, может также использоваться банк данных угроз безопасности информации ФСТЭК России [12]. В настоящее время в списке угроз банка содержатся 182 идентифицированные угрозы для различных типов развертывания информационных систем. В частности, согласно данному списку для грид-систем актуальными являются угрозы:

- УБИ.001 – автоматического распространения вредоносного кода в грид-системе;
- УБИ.002 – агрегирования данных, передаваемых в грид-системе;
- УБИ.047 – нарушения работоспособности грид-системы при нетипичной сетевой нагрузке;
- УБИ.081 – несанкционированного доступа к локальному компьютеру через клиента грид-системы;
- УБИ.110 – перегрузки грид-системы вычислительными заданиями;
- УБИ.147 – распространения несанкционированно повышенных прав на всю грид-систему.

Для суперкомпьютерных систем актуальными будут угрозы:

УБИ.029 – использования вычислительных ресурсов суперкомпьютера «паразитными» процессами;

УБИ.106 – отказа в обслуживании системой хранения данных суперкомпьютера;

УБИ.146 – прямого обращения к памяти вычислительного поля суперкомпьютера;

УБИ.161 – чрезмерного использования вычислительных ресурсов суперкомпьютера в ходе интенсивного обмена межпроцессорными сообщениями.

В базу данных уязвимостей банка включено около 13 000 идентифицированных уязвимостей, в том числе значительное количество уязвимостей, характерных только для грид-систем. В частности, УБИ 001 может быть реализована при условии наличия у нарушителя привилегий легального пользователя грид-системы, например, с использованием уязвимости 2014-00013 (табл. 1).

Таблица 1

Описание и основные параметры уязвимости 2014-00013

Параметр	Содержание
Описание уязвимости	Система обмена программными сообщениями Apache Qpid содержит уязвимость, связанную с ошибкой в механизме аутентификации во время обработки теневых подключений AMQP-клиентов. С помощью специально сформированного запроса злоумышленник может представиться законным пользователем
Вендор	Apache Software Foundation
Наименование ПО	Apache Qpid
Версия ПО	0.16, 0.14
Тип ПО	Сетевое ПО
Операционные системы и аппаратные платформы	Сообщество свободного программного обеспечения Linux.x64; сообщество свободного программного обеспечения Linux.x86; Microsoft Corp. Windows.x64; Microsoft Corp. Windows.x86
Тип ошибки	Неправильная аутентификация
Класс уязвимости	Уязвимость архитектуры
Дата выявления	22.06.2012
Базовый вектор уязвимости	AV:N/AC:L/Au:N/C:N/I:P/A:N
Уровень опасности уязвимости	Средний уровень опасности (базовая оценка CVSS составляет 5)
Возможные меры по устранению уязвимости	Обновление ПО до версии 0.17 или выше
Статус уязвимости	Подтверждена производителем
Наличие эксплойта	Существует
Информация об устранении	Уязвимость устранена
Идентификаторы других систем описаний уязвимостей	CVE: CVE-2012-3467 OSVDB: OSVDB ID:84562
Язык разработки ПО	Java, C++

Для примера оценим риск реализации угрозы автоматического распространения вредоносного кода в грид-системе (УБИ 001) посредством эксплуатации уязвимости 2014-00013 с использованием наиболее распространенной, востребованной и проверенной на практике системы оценки общеизвестных уязвимостей (common vulnerability scoring system, CVSS 2.0) [13–15], которая состоит из трех групп метрик (критериев).

Величина риска, соответствующая прогнозируемым среднегодовым потерям в результате инцидентов безопасности *ALE* (annual loss expectancy), определяется по формуле [15]

$$ALE = AV^* \cdot EF \cdot ARO, \quad (1)$$

где *AV\** (asset value) – стоимость подверженных риску активов грид-системы (данных, программ, аппаратуры и т. д.);

*EF* (exposure factor) – степень уязвимости актива к угрозе;

*ARO* (annualized rate of occurrence) – среднегодовая частота возникновения инцидентов (величина, представляющая собой ожидаемую частоту реализации угрозы в год).

На основании анализа основных параметров уязвимости 2014-00013 и конкретной среды эксплуатации грид-системы определим значения базовых, временных и контекстных метрик данной уязвимости, которые приведены в табл. 2 (выделены соответствующими цветами).

Таблица 2

Значения метрик уязвимости 2014-00013

Базовые метрики (BM)			Временные метрики (TM)					Контекстные метрики (EM)					
Вектор доступа (AV)			Возможность использования (E)					Возможность косвенного ущерба (CDP)					
L	A	N	ND	U	POC	F	H	ND	N	L	LM	MN	H
Сложность доступа (AC)			Уровень исправления (RL)					Плотность целей (TD)					
H	M	L	ND	OF	T	W	U	ND	N	L	M	H	
Аутентификация (AU)			Степень достоверности источника (RC)					Требования к конфид. (CR)					
M	S	N	ND	UC	UR	C	ND	L	M	H			
Влияние на конфид. (C)								Требования к целостности (IR)					
N	P	C						ND	L	M	H		
Влияние на целостность (I)								Требования к доступности (AR)					
N	P	C						ND	L	M	H		
Влияние на доступность (A)													
N	P	C											

*Базовые метрики (BM)*. Значения базовых метрик определены в базовом векторе уязвимости (AV:N/AC:L/Au:N/C:N/I:P/A:N) (см. табл. 1).

*Временные метрики (TM)*. С помощью метрики «возможность использования» (E) измеряется существующее состояние доступности кода или метода эксплуатации уязвимости. В табл. 1 указано, что эксплойт существует, поэтому согласно [14] E примет значение F (функциональная). Это означает, что функциональный код эксплойта доступен и применим в большинстве ситуаций, где существует уязвимость. Метрика «уровень исправления» (RL) примет

значение OF (официальное исправление), так как в табл. 1 официально информируется, что уязвимость устранена. Метрика «степень достоверности источника» (RC) примет значение C (подтверждено), так как в табл. 1 сообщается, что статус уязвимости подтвержден производителем.

*Контекстные метрики (EM).* Являются необязательными и отражают характеристики уязвимости, которые относятся к конкретной среде эксплуатации и характерны только для нее. Контекстные метрики определяются пользователями, поскольку они точнее могут оценить потенциальное воздействие уязвимости в рамках своей собственной среды. У нас нет полных сведений о конкретной используемой или проектируемой грид-системе, поэтому значения данных метрик будем задавать ориентировочно (на основании опыта, здравого смысла и общих сведений об аналогичных системах).

Метрике «возможность косвенного ущерба» (CDP) присвоим значение LM (от низкой до средней). Это означает, что может иметь место умеренная потеря дохода или производительности грид-системы. Метрике «плотность целей» (TD) присвоим значение L (низкое), т. е. риску подвержено от 1 до 25 % всех систем среды. Метрикам «требования к конфиденциальности» (CR) и «требования к доступности» (AR) присвоим значение ND (не определено), так как базовые метрики «влияние на конфиденциальность» (C) и «влияние на доступность» (A) принимают значение N (отсутствует). Метрике «требования к целостности» (IR) присвоим значение M (среднее), так как базовая метрика «влияние на целостность» (I) принимает значение P (частичное).

В результате получим суммарный вектор уязвимости:

$$AV:N/AC:L/Au:N/C:N/I:P/A:N/E:F/RL:OF/RC:C/CDP:LM/TD:L/CR:ND/IR:M/AR:ND.$$

Далее воспользуемся калькулятором CVSS [12] и получим контекстную оценку уязвимости environmental score (ES) = 1,5. Согласно шкале FortiGuard [16] уровень опасности данной уязвимости низкий. После нормирования получим [15]

$$EF = CVSS \text{ Rating} / 10 = 0,15. \quad (2)$$

Определяем среднегодовую частоту возникновения инцидентов:

$$ARO = P_E \cdot F_E, \quad (3)$$

где  $P_E$  – вероятность использования уязвимости;

$F_E$  – количество попыток в год, когда имелась принципиальная возможность осуществить эксплуатацию уязвимости.

Вероятность эксплуатации уязвимости вычисляется по формуле [15]

$$P_E = AV \cdot AC \cdot Au \cdot E \cdot RL \cdot RC = 1 \cdot 0,71 \cdot 0,704 \cdot 0,95 \cdot 0,87 \cdot 1 = 0,4. \quad (4)$$

Стоимость подвергаемых риску активов  $AV^*$  и прогнозируемых частот реализации конкретных угроз (инцидентов ИБ)  $F_E$  определяется с использованием аппарата теории вероятностей и математической статистики, аналитических методов, а также экспертных оценок. В разд. 3 предложен один из таких статистических методов, основанный на теоретических и экспериментальных исследованиях.

Приведем конкретный пример расчета величины риска. Предположим, что прибыль от работы некоторой условной коммерческой грид-системы составляет 750 млн руб. в месяц. В результате автоматического распространения вредоносного кода грид-система (или один из ее узлов) вышла из строя. На восстановление ее работоспособности планируется затратить одни сутки. Стоимость затрат на ремонт составит примерно 25 млн руб. По формуле (2) определяем степень уязвимости грид-системы к угрозе реализации нелегитимных действий нарушителем  $EF = 0,15$ . Вероятность осуществления нарушителем нелегитимных действий (например, по уровню квалификации) вычисляем по формуле (4) –  $P_E = 0,4$ . Количество попыток в год, когда нарушитель в принципе может осуществлять такие действия (например, посещает

объекты грид-системы), определяем из опыта или статистических данных –  $F_E = 10$ . Далее вычисляем  $AV^* = 750/30 + 25 = 50$ . По формуле (3) определяем  $ARO = 0,4 \cdot 10 = 4$ . Следовательно, в соответствии с формулой (1)

$$ALE = AV^* \cdot EF \cdot ARO = 50 \cdot 0,15 \cdot 4 = 30 \text{ млн руб.}$$

Аналогичным образом можно оценить риски реализации других угроз.

### 3. Статистические методы оценки информационных атак

Для анализа угроз на основе накопленных данных об инцидентах ИБ могут применяться различные статистические методы. Предлагается метод оценки инцидентов, которые связаны с реализацией возможных угроз, воздействующих на систему, на основе использования критерия  $\chi^2$  [17].

Предположим, что частоты инцидентов, связанных с реализацией угроз  $E_1, E_2, \dots, E_i, \dots, E_r$  за промежуток времени  $T_1$  (например, за текущий месяц, квартал, год), составляют  $\mu_1, \mu_2, \dots, \mu_i, \dots, \mu_r$ ; частоты инцидентов, связанных с реализацией тех же угроз  $E_1, E_2, \dots, E_i, \dots, E_r$  за промежуток времени  $T_2$  (например, за прошлый месяц, квартал, год), – соответственно  $\nu_1, \nu_2, \dots, \nu_i, \dots, \nu_r$ .

Используем критерий  $\chi^2$  для сравнения данных последовательностей частот на предмет их однородности. Выясним, не произошли ли изменения в распределении вероятностей реализации угроз во временном промежутке  $T_2$  по сравнению с временным промежутком  $T_1$ . Для этого вычислим статистику [17]:

$$\chi^2 = m \cdot n \sum_{i=1}^r \frac{1}{\mu_i + \nu_i} \left( \frac{\mu_i}{m} - \frac{\nu_i}{n} \right)^2, \quad (5)$$

где  $m = \sum_{i=1}^r \mu_i$ ,  $n = \sum_{i=1}^r \nu_i$ .

Данная статистика при больших объемах выборок  $m$  и  $n$  имеет распределение  $\chi^2$  с  $r-1$  степенями свободы. Проверяемая гипотеза об однородности отклоняется, когда вычисленное значение статистики  $\chi^2$  больше критического значения  $\chi_{\alpha}^2$ , или достигнутый уровень значимости ( $p$ -value) [17]

$$p = P(\chi^2 > \chi_p^2) = \frac{1}{2^{r-1} \Gamma\left(\frac{r-1}{2}\right)} \int_{\chi_p^2}^{\infty} x^{\frac{r-1}{2}-1} e^{-\frac{x}{2}} dx \quad (6)$$

меньше заданного уровня значимости (заданной вероятности ошибки первого рода)  $\alpha$ , где  $\Gamma\left(\frac{r-1}{2}\right)$  – гамма-функция Эйлера.

Проиллюстрируем применение предложенного метода на примере отчетов JSOC Security flash report, которые основаны на данных, полученных в коммерческом центре мониторинга и реагирования на инциденты ИБ JSOC (jet security operation center) [18]. Отчеты являются сводными материалами и результатами анализа инцидентов, выявленных командой JSOC как в рамках оказания регулярных услуг мониторинга и реагирования на инциденты, так и в качестве консультативно-аналитической поддержки компаний российского рынка. Они предназначены для информирования служб информационных технологий и ИБ об основных трендах, касающихся угроз ИБ.

Рассмотрим внутренние инциденты, т. е. инциденты, инициаторами и причиной которых становились действия внутренних сотрудников клиентов JSOC: халатность в соблюдении политик ИБ или их прямое нарушение, компрометация или передача учетных данных сотрудников, злонамеренные и незлонамеренные воздействия на бизнес-процессы и функционирование

систем клиентов. К внутренним пользователям – инициаторам инцидента – относятся все пользователи с возможностью доступа в локальную сеть без преодоления периметра, в том числе подрядчики и контрагенты. Сводные частоты данных инцидентов приведены в табл. 3.

Таблица 3

Статистика по внутренним инцидентам за третий и второй кварталы 2014 г.

Суммарные частоты инцидентов		Частоты инцидентов, связанных с реализацией типов угроз						
		$E_1$	$E_2$	$E_3$	$E_4$	$E_5$	$E_6$	$E_7$
		Утечка конфиденциальных данных	Нарушение политик доступа в Интернет	Несанкционированные активности в рамках удаленного доступа	Компрометация внутренних учетных записей	Нелегитимные изменения в ИТ-системах	Вирусные атаки, включая ransomware, zero-day	Нелегитимная работа с привилегир. учетными записями и прочие инциденты
3 кв.	$m$	$\mu_1$	$\mu_2$	$\mu_3$	$\mu_4$	$\mu_5$	$\mu_6$	$\mu_7$
	11 692	3 028	2 011	573	1 649	935	1 380	2 116
2 кв.	$n$	$\nu_1$	$\nu_2$	$\nu_3$	$\nu_4$	$\nu_5$	$\nu_6$	$\nu_7$
	10 891	2 962	1 775	468	1 579	915	1 232	1960

С помощью статистики  $\chi^2$  оценим, произошли ли изменения в распределении частот внутренних инцидентов в третьем квартале по сравнению со вторым кварталом. По формуле (5) получим значение  $\chi^2 = 13,72$  (для шести степеней свободы), которое не превышает критического значения  $\chi_{6,p}^2 = 16,81$ , соответствующего согласно формуле (6) уровню значимости  $p = 0,01$ . Это означает, что не произошло существенных изменений в распределении внутренних инцидентов в третьем квартале по сравнению со вторым.

Для сравнения рассмотрим внешние инциденты, т. е. инциденты, инициаторами и причиной которых становились действия лиц, не являющихся внутренними пользователями клиентов JSOC. Сводные частоты данных инцидентов приведены в табл. 4.

Таблица 4

Статистика по внешним инцидентам за третий и второй кварталы 2014 г.

Суммарные частоты инцидентов		Частоты инцидентов, связанных с реализацией типов угроз					
		$E_1$	$E_2$	$E_3$	$E_4$	$E_5$	$E_6$
		Атаки на веб-приложения	Brute-force и компрометация учетных данных внешних сервисов	Компрометация административных учетных записей	Атаки на управляющие протоколы систем	DDoS-атаки	Прочие внешние атаки
3 кв.	$m$	$\mu_1$	$\mu_2$	$\mu_3$	$\mu_4$	$\mu_5$	$\mu_6$
	7 166	2 723	1 835	100	373	215	1 920
2 кв.	$n$	$\nu_1$	$\nu_2$	$\nu_3$	$\nu_4$	$\nu_5$	$\nu_6$
	6 818	2 482	1 677	157	198	198	2 106

С помощью статистики  $\chi^2$  оценим, произошли ли изменения в распределении частот внешних инцидентов в третьем квартале по сравнению со вторым. По формуле (5) получим значение  $\chi^2 = 85,23$  (для пяти степеней свободы). Оно значительно превышает критическое

значение  $\chi^2_{5,p} = 15,09$ , которое согласно формуле (6) соответствует уровню значимости  $p = 0,01$ . Это означает, что отклонение от гипотезы об однородности распределений весьма значимо. Статистика показывает, что произошли существенные изменения в распределении внешних инцидентов. Отмечается смещение внешних атак на прикладной уровень. Одной из причин этого является существенное увеличение количества внешних ресурсов и веб-сервисов в компаниях практически всех отраслевых сегментов.

Предложенный метод носит универсальный характер и может использоваться для различных типов развертывания информационных систем (том числе для грид-систем) с целью выявления тенденций в динамике угроз. Действительно, если известны частоты инцидентов  $\mu_1, \mu_2, \dots, \mu_i, \dots, \mu_r$  за текущий месяц, квартал, полугодие и при сравнении их с частотами инцидентов  $\nu_1, \nu_2, \dots, \nu_i, \dots, \nu_r$  за прошлый год не выявлено существенного изменения в распределении вероятностей инцидентов (значение статистики  $\chi^2$  однородности не превышает заданного уровня), то с определенной вероятностью можно предположить, что эти показатели будут стабильны в среднесрочной перспективе. Это позволит спрогнозировать потенциальные частоты инцидентов  $\eta_1, \eta_2, \dots, \eta_i, \dots, \eta_r$  на текущий год, которые, в свою очередь, могут быть использованы при оценке рисков ИБ.

Если же значение статистики  $\chi^2$  однородности значительно превышает заданный уровень, то необходимо выявлять причины сложившейся ситуации с изменением распределения инцидентов и проводить углубленный анализ угроз, воздействующих на систему.

Предлагаемый подход к анализу угроз ИБ позволяет провести оценку защищенности грид-системы, функционирующей в условиях воздействия рассматриваемого класса угроз, а также эффективности комплекса мероприятий и средств противодействия данным угрозам.

### Заключение

На основании анализа опыта разработки распределенных вычислительных систем можно сделать вывод, что методы управления контролем безопасного доступа к услугам грид-систем должны обладать возможностями:

- осуществления однократной аутентификации пользователей;
- делегирования прав;
- интеграции с локальными системами безопасности и обмена политиками безопасности;
- обнаружения (предотвращения) угроз, вторжений (компьютерных атак), направленных на преднамеренный несанкционированный доступ к информации.

Угрозы безопасности информации в грид-системе определяются по результатам оценки и анализа возможных уязвимостей, способов реализации угроз безопасности информации и последствий от нарушения безопасности информации. Для оценки рисков ИБ рекомендуется использовать общедоступную базу данных CVE, банк данных ФСТЭК России (в особенности для государственных грид-систем), а также систему оценки общеизвестных уязвимостей CVSS и статистические критерии, в частности критерий  $\chi^2$ .

Приоритетным направлением совершенствования средств и систем ИБ грид-систем является постепенный переход к созданию многоагентных систем защиты информации, реализующих централизованно-децентрализованный принцип построения, а также применение современных способов разграничения доступа к информации, в том числе на основе мультибиометрической аутентификации пользователей [10].

### Список литературы

1. Информационная технология. Модель открытой грид-системы. Основные положения : ГОСТ Р 55768–2013. – М. : Стандартинформ, 2014. – 78 с.
2. О проблеме защиты информационных и вычислительных ресурсов в распределенных вычислительных сетях типа грид / М.О. Калинин [и др.] // XIV Санкт-Петербургская Междунар. конф. «Региональная информатика (РИ–2014)» : материалы конф. – СПб., 2014. – С. 553–554.

3. Information technology – Security techniques – Guidelines for the use and management of trusted third party services. ITU-T Recommendation X.842 (ISO/IEC TR 14516). – Montreal, 2000. – 35 p.
4. Сабанов, А.Г. О неизвлекаемости закрытых ключей / А.Г. Сабанов // Защита информации. Инсайд. – 2015. – № 2. – С. 30–33.
5. Информационные технологии и безопасность. Критерии оценки безопасности информационных технологий. Часть 3. Гарантийные требования безопасности : СТБ 34.101.3–2014 (ISO/IEC 15408-3:2008). – Минск : Госстандарт, 2014. – 132 с.
6. Андреев, Д.С. Ритм набора парольной фразы как дополнительный признак аутентификации пользователя / Д.С. Андреев // Сб. науч. тр. IV Междунар. науч.-практ. конф. – Курск, 2014. – С. 68–74.
7. Старовойтов, В.В. Биометрические системы контроля доступа по отпечаткам пальцев / В.В. Старовойтов // Информатика. – 2015. – № 1. – С. 26–38.
8. Приказ Оперативно-аналитического центра при Президенте Республики Беларусь от 16.01.2015 № 3 «О внесении дополнений и изменений в приказ Оперативно-аналитического центра при Президенте Республики Беларусь от 30.08.2013 № 62» [Электронный ресурс]. – 2015. – Режим доступа : [http://www.oac.gov.by/files/files/pravo/prikazi\\_oac/Prikaz\\_OAC\\_3.htm](http://www.oac.gov.by/files/files/pravo/prikazi_oac/Prikaz_OAC_3.htm). – Дата доступа : 03.10.2015.
9. Методический документ. Меры защиты информации в государственных информационных системах (утвержден ФСТЭК России 11.02.2014) [Электронный ресурс]. – 2015. – Режим доступа : <http://fstec.ru/tekhnicheskaya-zashchita-informatsii/dokumenty/114-spetsialnye-normativnyue-dokumenty/805-metodicheskij-dokument>. – Дата доступа : 03.10.2015.
10. Язов, Ю.К. Развитие информационных технологий и приоритетные направления исследований по совершенствованию защиты общих информационных ресурсов Беларуси и России в период 2016–2021 годов / Ю.К. Язов // Комплексная защита информации : материалы XX науч.-практ. конф., Минск, 19–21 мая 2015 г. – Минск : РИВШ, 2015. – С. 12–16.
11. Common Vulnerabilities and Exposures [Electronic resource]. – 2015. – Mode of access : <http://www.cve.mitre.org>. – Date of access : 30.07.2015.
12. Банк данных угроз безопасности информации Федеральной службы по техническому и экспортному контролю (ФСТЭК России) [Электронный ресурс]. – 2015. – Режим доступа : <http://bdu.fstec.ru>. – Дата доступа : 03.10.2015.
13. NVD Common Vulnerability Scoring System Support v.2 [Electronic resource]. – 2015. – Mode of access : <http://www.nvd.nist.gov/cvss.cfm?calculator&version=2>. – Date of access : 03.10.2015.
14. Система оценки общеизвестных уязвимостей. Рекомендация. МСЭ-Т X.1521 (04/2011) [Электронный ресурс]. – 2015. – Режим доступа : [https://www.itu.int/rec/dologin\\_pub.asp?lang=e&id=T-REC-X.1521-201104-I!!PDF-R&type=items](https://www.itu.int/rec/dologin_pub.asp?lang=e&id=T-REC-X.1521-201104-I!!PDF-R&type=items). – Дата доступа : 03.10.2015.
15. Трубей, А.И. Оценка рисков информационной безопасности с использованием существующей нормативной и методической базы / А.И. Трубей // Информатика. – 2015. – № 2. – С. 102–114.
16. Vulnerability Severity Level [Electronic resource]. – 2015. – Mode of access : <http://www.fortiguard.com/static/intrusionprevention.html>. – Date of access : 03.10.2015.
17. Крамер, Г. Математические методы статистики / Г. Крамер. – М. : Мир, 1976. – 648 с.
18. JSOC Security flash report 2014 Q3 [Электронный ресурс]. – 2015. – Режим доступа : <http://www.jet.msk.su/upload/content/jsoc-security-flash-report-q3-v-3.pdf>. – Дата доступа : 03.10.2015.

Поступила 04.11.2015

*Объединенный институт проблем  
информатики НАН Беларуси,  
Минск, Сурганова, 6  
e-mail: trubeia@newman.bas-net.by*

**A.I. Trubei**

**ANALYSIS OF GRID-SYSTEMS INFORMATION SECURITY ASPECTS**

The article presents the analysis of Grid-systems information security aspects, Also, risk assessment methods, using public databases, expert systems and statistical techniques are proposed.

**XV МЕЖДУНАРОДНАЯ КОНФЕРЕНЦИЯ  
«РАЗВИТИЕ ИНФОРМАТИЗАЦИИ И ГОСУДАРСТВЕННОЙ СИСТЕМЫ  
НАУЧНО-ТЕХНИЧЕСКОЙ ИНФОРМАЦИИ» (РИНТИ–2016)  
17 ноября 2016 г., Минск, Беларусь**

**ЦЕЛЬ КОНФЕРЕНЦИИ**

Анализ состояния и проблем развития информатизации и государственной системы научно-технической информации в Республике Беларусь; разработка новых подходов к научно-методическому обеспечению развития информатизации, созданию информационных технологий и информационной инфраструктуры Беларуси, построению информационного общества, реализации Стратегии развития информатизации в Республике Беларусь на 2016–2022 гг. и государственной программы развития цифровой экономики и информационного общества на 2016–2022 гг.

**ТЕМАТИКА КОНФЕРЕНЦИИ**

- Анализ состояния и проблем развития единого информационного пространства, методы и средства обеспечения его формирования, использования и развития в интересах граждан, общества и государства.
- Технологии информационного общества: электронное государство, электронное правительство, электронное обучение, электронная медицина, электронные библиотеки, криптографические методы защиты информации и др.
- Разработка и внедрение автоматизированных информационных систем управления и систем научно-технической информации, реализация проектов государственных программ в сфере информатизации, создание и внедрение автоматизированных библиотечно-информационных систем и технологий.
- Интерактивные сервисы, реализующие обслуживание населения по принципу «одно окно», реализация инфраструктуры электронной цифровой подписи и документооборота.
- Проведение научно-технической экспертизы; оценка научно-технического уровня и новизны проектов и программ информатизации, научной и практической значимости предполагаемых результатов их выполнения.
- Формирование интегрированной системы стандартизации (унификации), сертификации, экспертизы, оценки эффективности реализации программ и проектов в области информатизации.
- Научно-информационные компьютерные сети, технологии высокоскоростного информационного обмена между организациями с доступом в международные научно-образовательные сети.
- Корпоративная сеть библиотек Беларуси и ее информационные ресурсы, автоматизация Национальной библиотеки Беларуси, генерация собственных информационных ресурсов, обеспечение доступа к национальным и зарубежным библиотечным информационным ресурсам.
- Региональная информатизация и информационная инфраструктура.

**РАБОЧИЕ ЯЗЫКИ КОНФЕРЕНЦИИ**

В качестве рабочих языков устанавливаются белорусский, русский и английский.

**ПОРЯДОК ПРЕДСТАВЛЕНИЯ И ТРЕБОВАНИЯ К ОФОРМЛЕНИЮ ДОКЛАДОВ**

Доклады в электронном виде представляются на русском, белорусском или английском языках по электронной почте [vengerov@basnet.by](mailto:vengerov@basnet.by) (в случае возникновения проблем при доставке – на адреса [gsnti@basnet.by](mailto:gsnti@basnet.by) или [griganec@bas-net.by](mailto:griganec@bas-net.by)). Белорусские авторы предоставляют

также в оргкомитет конференции *в обязательном порядке* акт экспертизы о возможности опубликования доклада в открытой печати (нарочным или письмом на почтовый адрес конференции).

Доклады будут опубликованы в сборнике докладов и предоставлены всем докладчикам в день открытия конференции.

Участие в конференции – бесплатное.

Авторам докладов необходимо зарегистрироваться на сайте конференции и указать полностью фамилию, имя и отчество, организацию, должность, ученую степень и ученое звание, e-mail, телефоны (мобильный, служебный или домашний). Эта информация (кроме телефонов) будет размещена в сборнике докладов конференции в разделе НАШИ АВТОРЫ.

Объем представляемых докладов – до шести нумерованных страниц формата А4.

Доклад должен включать: УДК (в левом верхнем углу); название доклада; И.О.Ф. авторов; полное название организации, города, страны; аннотацию (не более 100 слов); текст доклада (с обязательными ссылками на литературу и рисунки); список литературы (если имеется), оформленный согласно приложению 2 к инструкции ВАК РБ по оформлению диссертации (<http://www.vak.org.by>).

Поля: верхнее – 2,5 см, нижнее – 3,2 см, левое – 2,5 см, правое – 2,5 см, абзацный отступ – 1,0 см. Интервал между строками и абзацами – одинарный. Текст должен входить в рамку 24×16 см и набираться с переносами. При подготовке текста доклада используются шрифты семейства Times New Roman Суг следующих размеров и типов:

УДК – прописные 12 пт; **НАЗВАНИЕ ДОКЛАДА** – прописные полужирные 12 пт; И.О.Ф. – строчные 12 пт; страна, город – строчные 12 пт; *аннотация* – 10 пт курсив (отступ слева и справа – 1 см); **заголовки разделов** – строчные полужирные 12 пт; **заголовки подразделов** – строчные полужирные 12 пт; основной текст – 12 пт без нумерации страниц.

Основной текст может включать **Введение** (тогда обязательно и **Заключение**). Все заголовки должны быть выровнены по левому краю. В возрастающем порядке должны быть указаны ссылки на литературу, рисунки и таблицы (черно-белые, четкие), если такие имеются.

Образцы оформления докладов представлены также на сайте конференции.

Тексты докладов проходят научную и редакторскую обработку, которая согласовывается с авторами.

## АДРЕС КОНФЕРЕНЦИИ

220012, Республика Беларусь, Минск, ул. Сурганова, 6, ОИПИ НАН Беларуси  
(<http://www.uiip.basnet.by>, тел.+375 17 284-25-26).

### Председатель программного комитета:

Тузиков Александр Васильевич +375 17 284-21-75 tuzikov@newman.bas-net.by

### Заместители председателя программного комитета:

Алюшкевич Виктор Бернардович +375 17 284-09-85 aliushkevich@newman.bas-net.by

Григянец Ромуальд Брониславович +375 17 284-25-26 griganec@bas-net.by

### Ответственные секретари:

Венгеров Виктор Николаевич +375 17 284-12-08 vengerov@basnet.by

Мисякова Галина Тихоновна +375 17 217-74-06 gsnti@basnet.by

**Е-mail:** vengerov@basnet.by, gsnti@basnet.by

**Сайт:** <http://lirs.basnet.by/opacpage/rinti>

## ПРАВИЛА ДЛЯ АВТОРОВ

1. Статьи принимаются в редакцию через электронную систему подачи по адресу <http://jinfo.bas-net.by> в формате файлов текстовых редакторов Microsoft Word 97 и Word 2000 для Windows. Основной текст статьи набирается с переносами шрифтом Times New Roman 11 пт, интервал между строками – одинарный, абзацный отступ 1 см, поля по 2,5 см со всех сторон.

2. Статья должна иметь индекс УДК (универсальная десятичная классификация).

3. Название статьи, фамилии всех авторов и аннотация должны быть переведены на английский язык. Для каждого из авторов приводится развернутое название учреждения с полным почтовым адресом, а также номер телефона и электронный адрес (e-mail) для связи с редакцией.

4. Формулы, иллюстрации, таблицы, встречающиеся в статье, должны быть пронумерованы в соответствии с порядком цитирования в тексте. Ссылки на рисунки и таблицы в тексте обязательны. Необходимо избегать повторения одних и тех же данных в таблицах, графиках и тексте статьи.

Рисунки должны быть выполнены с хорошим разрешением в масштабе, позволяющем четко различать надписи и обозначения. Подрисовочные подписи с расшифровкой всех позиций, представленных на рисунке, набираются шрифтом гарнитуры основного текста, размер символов 9 пт. Цветные иллюстрации печатаются только в том случае, когда это необходимо для понимания излагаемого материала.

5. Набор формул выполняется в формульных редакторах Microsoft Equation или Math Type и должен быть единообразным по применению шрифтов и знаков по всей статье.

Прямо ( ) набираются: греческие и русские буквы; математические символы ( $\sin$ ,  $\lg$ ,  $\infty$ ); символы химических элементов (C, Cl,  $\text{CHCl}_3$ ); цифры (римские и арабские); векторы; индексы (верхние и нижние), являющиеся сокращениями слов.

Курсивом (–) набираются: латинские буквы – переменные, символы физических величин (в том числе и в индексе).

6. Сокращения в тексте статьи (за исключением единиц измерения) могут быть использованы только после упоминания полного термина. Единицы измерения физических величин следует приводить в Международной системе СИ.

7. Литература приводится автором общим списком в конце статьи. Ссылки на литературу в тексте идут по порядку и обозначаются цифрой в квадратных скобках. Ссылаться на неопубликованные работы не допускается. С примерами оформления библиографического описания в списке литературы можно ознакомиться в приложении 2 к *Инструкции по оформлению диссертации, автореферата и публикаций по теме диссертации* на сайте Высшей аттестационной комиссии Республики Беларусь <http://vak.org.by>.

8. Поступившие в редакцию статьи направляются на рецензирование специалистам. Основным критерием целесообразности публикации является новизна и информативность статьи. Если по рекомендациям рецензента статья возвращается автору на доработку, а переработанная рукопись вновь рассматривается редколлегией, датой поступления считается день получения редакцией ее окончательного варианта. Статьи не по профилю журнала возвращаются авторам после заключения редколлегии.

9. Статьи, направляемые на доработку, должны быть возвращены в исправленном виде с ответами на все вопросы.

10. Редакция журнала предоставляет возможность первоочередного опубликования статей, представленных лицами, которые осуществляют послевузовское обучение (аспирантура, докторантура, соискательство) в год завершения обучения.

11. Авторы несут ответственность за направление в редакцию статей, уже опубликованных ранее, или статей, принятых к публикации другими изданиями.

12. Редакция оставляет за собой право на редакционные изменения, не искажающие основное содержание статьи.

***Журнал «Информатика» включен Высшей аттестационной комиссией Республики Беларусь в список научных изданий для опубликования результатов диссертационных исследований.***

## Индексы

**00827**

для индивидуальных  
подписчиков

**008272**

для предприятий и  
организаций