

ОБЪЕДИНЕННЫЙ ИНСТИТУТ ПРОБЛЕМ ИНФОРМАТИКИ
НАЦИОНАЛЬНОЙ АКАДЕМИИ НАУК БЕЛАРУСИ

ИНФОРМАТИКА

ЕЖЕКВАРТАЛЬНЫЙ НАУЧНЫЙ ЖУРНАЛ

Издается с января 2004 г.

№ 1(37) • январь-март 2013

СОДЕРЖАНИЕ

МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ

- Ерофеев В.Т., Бондаренко В.Ф.** Преобразование пучков электромагнитных волн при прохождении через экран из кирального метаматериала 5
- Буза М.К., Горшунов В.С.** Система моделирования поведения клиентов организаций... 18
- Баркалин В.В., Долгая Я.В., Козлова О.А.** Моделирование предполимеризационного комплекса полимеров с молекулярными отпечатками три-о-ацетиладенозина 26
- Родченко В.Г., Олизарович Е.В., Жукевич А.И.** Применение методов распознавания для построения систем диагностики компьютерных сетей 36

ОБРАБОТКА ИЗОБРАЖЕНИЙ И РАСПОЗНАВАНИЕ ОБРАЗОВ

- Борискевич А.А., Руис Л.А.** Гибридный алгоритм сжатия изображений без потерь на основе дискретного вейвлет-преобразования с адаптивным предсказанием..... 48
- Прадун Д.В.** Снижение размерности обучающих выборок при распознавании образов на космических изображениях с помощью метода главных компонент 57
- Сейеди С.А., Садыхов Р.Х.** Сравнение методов графической стеганографии 66

ПРИКЛАДНЫЕ ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

Лялинский А.А. Особенности построения прикладных программ с веб-доступом.....	76
Михута И.Ю. Алгоритм оценки и прогнозирования психофизической готовности и пригодности учащейся молодежи к военно-профессиональной деятельности	85

ПАРАЛЛЕЛЬНЫЕ ВЫЧИСЛЕНИЯ

Коваленко Н.С., Венгеров В.Н., Метельский В.М. Минимизация числа обрабатываемых устройств при распределенной организации вычислений	94
Коляда А.А., Чернявский А.Ф. Интегрально-характеристическая база модулярных систем счисления	106

ЛОГИЧЕСКОЕ ПРОЕКТИРОВАНИЕ

Поттосин Ю.В., Кардаш С.Н. Повышение быстродействия комбинационной схемы путем конвейеризации	120
--	-----

КОНФЕРЕНЦИИ

X Международная конференция «Computer data analysis & modeling» (CDAM-2013)	129
--	------------

Редактор Г.Б. Гончаренко
Корректор А.А. Михайлова
Компьютерная верстка В.И. Голенкевич

Сдано в набор 14.01.2013. Подписано в печать 01.03.2013.
Формат 60×84 1/8. Бумага офсетная. Гарнитура Таймс.
Усл. печ. л. 14,9. Уч.-изд. л. 14,6. Тираж 100 экз. Заказ 2.

Государственное научное учреждение «Объединенный институт проблем информатики Национальной академии наук Беларуси».
ЛИ № 02330/0549421 от 08.04.2009.
Ул. Сурганова, 6, 220012, Минск.

Отпечатано с оригинала-макета на ризографе Объединенного института проблем информатики Национальной академии наук Беларуси.
Ул. Сурганова, 6, 220012, Минск.

THE UNITED INSTITUTE OF INFORMATICS PROBLEMS
OF THE NATIONAL ACADEMY OF SCIENCES OF BELARUS

INFORMATICS

PUBLISHED QUATERLY

Issued since 2004

№ 1(37) • January-March 2013

CONTENTS

MATHEMATICAL MODELING

- Erofeenko V.T., Bondarenko V.F.** Transformation of beams of electromagnetic waves passing through a chiral metamaterial screen 5
- Bouza M.K., Gorshunov V.S.** Modeling organization clients behavior 18
- Barkaline V.V., Douhaya Y.V., Kozlova O.A.** QM/MM modeling of TAA imprinted polymers pre-polymerization complex 26
- Rodchenko V.G., Olizarovich E.V., Zhukevich A.I.** Pattern recognition methods for building a network diagnostic system 36

IMAGE PROCESSING AND PATTERN RECOGNITION

- Boriskevich A.A., Ruiz L.A.** Hybrid no-loss image compression algorithm based on discrete wavelet transformation with adaptive prediction 48
- Pradun D.V.** Reduction of training samples dimension in pattern recognition of space images using principal components analysis 57
- Seyyedi S.A., Sadykhov R.Kh.** Comparison of digital image steganography methods 66

APPLIED INFORMATION TECHNOLOGIES

Lyalinsky A.A. Website execution of cad modules	76
Mikhuta I.Y. Algorithm of the assessment and forecasting of psychophysical readiness and suitability of studying youth to military professional activity	85

PARALLEL COMPUTING

Kovalenko N.S., Vengherov V.N., Metelyskij V.M. Minimizing the number of processing units in distributed computations	94
Kolyada A.A., Chernyavsky A.F. Integrated characteristic base of modular number systems	106

LOGICAL DESIGN

Pottosin Yu.V., Kardash S.N. Increasing combinational circuit performance via pipelining.....	120
--	-----

CONFERENCES

X International conference «Computer data analysis & modeling» (CDAM-2013)	129
---	-----

МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ

УДК 517.958:537.8

В.Т. Ерофеев¹, В.Ф. Бондаренко²ПРЕОБРАЗОВАНИЕ ПУЧКОВ ЭЛЕКТРОМАГНИТНЫХ ВОЛН
ПРИ ПРОХОЖДЕНИИ ЧЕРЕЗ ЭКРАН ИЗ КИРАЛЬНОГО МЕТАМАТЕРИАЛА

Решается краевая задача проникновения осесимметричного пучка высокочастотных электромагнитных волн Лагерра – Гаусса через биизотропный плоский экран при ортогональном падении. Поле, прошедшее через экран, представляется аналитически в интегральном виде. Разрабатывается компьютерная программа для численного исследования структуры пучка в зависимости от материальных параметров экрана. Исследуются свойства фокусировки пучка волн экраном из кирального метаматериала.

Введение

В научной литературе активно исследуются процессы взаимодействия электромагнитных волн различных типов с материальными структурами, выполненными из композитных материалов [1]: киральных и биизотропных [2], бианизотропных [3], метаматериалов [4–6] и др. Важнейшими для приложений техническими структурами являются электромагнитные плоскостойкие экраны из композитных материалов, используемые для подавления внешних электромагнитных излучений. Также применяются слоистые покрытия технических устройств различного назначения для решения проблем электромагнитной совместимости технических систем и снижения радиозаметности объектов. Типовыми полями внешних источников являются плоские электромагнитные поля [7, 8], поля точечных источников [9], пучки электромагнитных волн [3], которые воздействуют на биизотропные экраны. Актуальными для приложений являются исследования гауссовых пучков волн в различных средах для оптического диапазона частот [10–16]. В книге [16] для исследования круговых пучков используется приближенный метод параболических уравнений, а в работе [10] исследованы отражательные свойства плоских пучков волн. В работах [11, 12] рассмотрены сингулярные пучки со сложной осевой симметрией при взаимодействии с полупространством резонансных сред. Значительный интерес для исследований представляют свойства фокусировки электромагнитных волн с помощью плоских экранов из метаматериалов [6, 9, 13].

В настоящей работе получено точное решение краевой задачи проникновения круговых пучков электромагнитных волн Лагерра – Гаусса через плоский слой из биизотропного материала. Пучок, прошедший через экран, представлен в интегральном виде. Разработана компьютерная программа для численного исследования геометрической структуры пучка в полупространстве по другую сторону слоя. Представлен графический материал для пучков волн в случае экрана из кирального метаматериала.

1. Постановка задачи

В пространстве R^3 с электрической и магнитной постоянными ϵ_0 , μ_0 размещен экран $D(0 < z < \Delta)$ толщины Δ . Экран D ограничен плоскостями $\Gamma_1(z=0)$, $\Gamma_2(z=\Delta)$ и заполнен биизотропным материалом с электромагнитными комплексными параметрами $\epsilon = \epsilon_r \epsilon_0$, $\mu = \mu_r \mu_0$, $G = G_r \sqrt{\epsilon_0 \mu_0}$, $Z = Z_r \sqrt{\epsilon_0 \mu_0}$. В пространстве $D_1(z < 0)$ ортогонально слою D распространяется круговой пучок электромагнитных волн Лагерра – Гаусса \vec{E}_0, \vec{H}_0 . Представим пучок в интегральном виде через базисные цилиндрические поля [17]:

$$\begin{aligned}\vec{E}_0 &= \int_0^\infty \left[a(\lambda) M_0^{(-1)}(\rho, z, \varphi; \lambda, k_0) + b(\lambda) M_0^{(-2)}(\rho, z, \varphi; \lambda, k_0) \right] d\lambda; \\ \vec{H}_0 &= h_0 \int_0^\infty \left[a(\lambda) M_0^{(-2)}(\rho, z, \varphi; \lambda, k_0) + b(\lambda) M_0^{(-1)}(\rho, z, \varphi; \lambda, k_0) \right] d\lambda,\end{aligned}\quad (1)$$

где $h_0 = \frac{1}{iZ_0}$, $Z_0 = \sqrt{\mu_0/\varepsilon_0}$, $k_0 = \frac{\omega}{c}$, $\omega = 2\pi f$ – круговая частота поля, c – скорость света,

$$a(\lambda) = Ag_n \left(\frac{\lambda}{\alpha_1} \right) \exp \left(-v_0(\lambda)h - \frac{1}{2} \left(\frac{\lambda}{\alpha_1} \right)^2 \right), \quad n = 0, 1, 2, \dots,$$

$$b(\lambda) = \frac{B}{v_0(\lambda)} g_n \left(\frac{\lambda}{\alpha_2} \right) \exp \left(-v_0(\lambda)h - \frac{1}{2} \left(\frac{\lambda}{\alpha_2} \right)^2 \right),$$

$$g_n(x) = (x^2 - 2n)L_n(x^2) + 2nL_{n-1}(x^2),$$

$L_n(z)$ – полиномы Лагерра,

$$v_0(\lambda) = \sqrt{\lambda^2 - k_0^2} \quad \text{при } \lambda \geq k_0, \quad v_0(\lambda) = -i\sqrt{k_0^2 - \lambda^2} \quad \text{при } k_0 > \lambda, \quad 0 \leq \lambda < \infty;$$

A, B – произвольные комплексные постоянные, $h > 0$, $\alpha_1 > 0$, $\alpha_2 > 0$; ρ, z, φ – цилиндрические координаты.

Компоненты первичного электрического поля (1) в сечении $z = -h$ определяются простыми аналитическими формулами

$$\begin{aligned}E_{0\varphi}|_{z=-h} &= (-1)^n \alpha_1^2 \rho A L_n \left((\alpha_1 \rho)^2 \right) \exp \left(-\frac{1}{2} (\alpha_1 \rho)^2 \right); \\ E_{0\rho}|_{z=-h} &= (-1)^n \alpha_2^2 \rho B L_n \left((\alpha_2 \rho)^2 \right) \exp \left(-\frac{1}{2} (\alpha_2 \rho)^2 \right).\end{aligned}\quad (2)$$

Обозначим поля: \vec{E}'_1, \vec{H}'_1 – отраженное поле в области D_1 ; \vec{E}_2, \vec{H}_2 – поле, проникшее в область D_2 ; $\vec{E}_1 = \vec{E}_0 + \vec{E}'_1$, $\vec{H}_1 = \vec{H}_0 + \vec{H}'_1$ – суммарное поле в D_1 ; \vec{E}, \vec{H} – поле в слое D .

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

Краевая задача 1. Требуется при заданном поле источника \vec{E}_0, \vec{H}_0 определить поля $\vec{E}'_1, \vec{H}'_1 \in C^1(D_1) \cap C(\bar{D}_1)$, $\vec{E}_2, \vec{H}_2 \in C^1(D_2) \cap C(\bar{D}_2)$, $\vec{E}, \vec{H} \in C^1(D) \cap C(\bar{D})$, для которых выполнены уравнения

$$\operatorname{rot} \vec{E} = i\omega(\mu \vec{H} + Z\vec{E}), \quad \operatorname{rot} \vec{H} = -i\omega(\varepsilon \vec{E} + G\vec{H}) \quad \text{в } D; \quad (3)$$

$$\operatorname{rot} \vec{E}_j = i\omega\mu_0 \vec{H}_j, \quad \operatorname{rot} \vec{H}_j = -i\omega\varepsilon_0 \vec{E}_j \quad \text{в } D_j, \quad (4)$$

граничные условия непрерывности тангенциальных составляющих полей на плоскостях Γ_1 и Γ_2

$$\left(\vec{E}_{1r} - \vec{E}_r\right)_{\Gamma_1} = 0, \quad \left(\vec{H}_{1r} - \vec{H}_r\right)_{\Gamma_1} = 0, \quad \left(\vec{E}_{2r} - \vec{E}_r\right)_{\Gamma_2} = 0, \quad \left(\vec{H}_{2r} - \vec{H}_r\right)_{\Gamma_2} = 0 \quad (5)$$

и условия излучения на бесконечности

$$\lim_{r \rightarrow \infty} r \left(\frac{\partial \vec{E}'_1}{\partial r} - ik_0 \vec{E}'_1 \right) = 0, \quad z < 0; \quad \lim_{r \rightarrow \infty} r \left(\frac{\partial \vec{E}_2}{\partial r} - ik_0 \vec{E}_2 \right) = 0, \quad z > \Delta, \quad (6)$$

где r – радиальная координата.

2. Представление электромагнитных полей

Для аналитического решения задачи (3) – (6) представим электромагнитные поля в областях D_j через базисные поля в интегральном виде:

$$\vec{E}_2 = \int_0^\infty [x_2(\lambda) M_0^{(-1)} + y_2(\lambda) M_0^{(-2)}] d\lambda, \quad \vec{H}_2 = h_0 \int_0^\infty [x_2(\lambda) M_0^{(-2)} + y_2(\lambda) M_0^{(-1)}] d\lambda, \quad z > \Delta; \quad (7)$$

$$\vec{E}'_1 = \int_0^\infty [x_1(\lambda) M_0^{(+1)} + y_1(\lambda) M_0^{(+2)}] d\lambda, \quad \vec{H}'_1 = h_0 \int_0^\infty [x_1(\lambda) M_0^{(+2)} + y_1(\lambda) M_0^{(+1)}] d\lambda, \quad z < 0, \quad (8)$$

где $x_j(\lambda)$, $y_j(\lambda)$ – функции, подлежащие определению.

Базисные цилиндрические поля определяются формулами [17, с. 131]

$$\vec{M}_m^{(\mp 1)}(\rho, z, \varphi; \lambda, k_0) = \vec{V}_m^{(1)}(\lambda \rho) \exp(\mp v_0(\lambda) z + im\varphi); \quad (9)$$

$$\vec{M}_m^{(\mp 2)}(\rho, z, \varphi; \lambda, k_0) = \frac{1}{k_0} (\mp v_0(\lambda) \vec{V}_m^{(2)}(\lambda \rho) + \lambda J_m(\lambda \rho) \vec{e}_z) \exp(\mp v_0(\lambda) z + im\varphi),$$

где
$$\vec{V}_m^{(1)}(\lambda \rho) = \frac{im}{\lambda \rho} J_m(\lambda \rho) \vec{e}_\rho - J'_m(\lambda \rho) \vec{e}_\varphi, \quad \vec{V}_m^{(2)}(\lambda \rho) = J'_m(\lambda \rho) \vec{e}_\rho + \frac{im}{\lambda \rho} J_m(\lambda \rho) \vec{e}_\varphi,$$

$v_0(\lambda) = \sqrt{\lambda^2 - k_0^2}$ при $\lambda \geq k_0$, $v_0(\lambda) = -i\sqrt{k_0^2 - \lambda^2}$ при $k_0 > \lambda$, $\vec{e}_\rho, \vec{e}_\varphi, \vec{e}_z$ – орты цилиндрической системы координат, $J_m(\cdot)$ – функции Бесселя.

В частном случае при $m = 0$ поля (9) имеют вид

$$\begin{aligned} \vec{M}_0^{(\mp 1)} &= \vec{V}_0^{(1)} \exp(\mp v_0 z), \quad \vec{M}_0^{(\mp 2)} = (\mp \bar{v}_0 \vec{V}_0^{(2)} + \bar{\lambda} J_0(\lambda \rho) \vec{e}_z) \exp(\mp v_0 z), \\ \vec{V}_0^{(1)} &= J_1(\lambda \rho) \vec{e}_\varphi, \quad \vec{V}_0^{(2)} = -J_1(\lambda \rho) \vec{e}_\rho, \quad \bar{v}_0(\lambda) = v_0(\lambda)/k_0, \quad \bar{\lambda} = \lambda/k_0. \end{aligned} \quad (10)$$

Для удовлетворения граничным условиям (5) вычислим тангенциальные составляющие $\vec{E}_{jr} = [\vec{e}_z, [\vec{E}_j, \vec{e}_z]]$ полей (1), (7), (8) на плоскостях Γ_1 и Γ_2 в базисе (10):

$$\vec{E}'_{1r} \Big|_{z=0} = \int_0^\infty [\hat{E}'_{1v_1}(\lambda) \vec{V}_0^{(1)}(\lambda \rho) + \hat{E}'_{1v_2}(\lambda) \vec{V}_0^{(2)}(\lambda \rho)] d\lambda,$$

$$\begin{aligned}\vec{H}'_{1\tau}\Big|_{z=0} &= \int_0^\infty \left[\hat{H}'_{1v_1}(\lambda) \vec{V}_0^{(1)}(\lambda\rho) + \hat{H}'_{1v_2}(\lambda) \vec{V}_0^{(2)}(\lambda\rho) \right] d\lambda; \\ \vec{E}_{2\tau}\Big|_{z=\Delta} &= \int_0^\infty \left[\hat{E}_{2v_1}(\lambda) \vec{V}_0^{(1)}(\lambda\rho) + \hat{E}_{2v_2}(\lambda) \vec{V}_0^{(2)}(\lambda\rho) \right] d\lambda,\end{aligned}\tag{11}$$

$$\begin{aligned}\vec{H}_{2\tau}\Big|_{z=\Delta} &= \int_0^\infty \left[\hat{H}_{2v_1}(\lambda) \vec{V}_0^{(1)}(\lambda\rho) + \hat{H}_{2v_2}(\lambda) \vec{V}_0^{(2)}(\lambda\rho) \right] d\lambda; \\ \vec{E}_{0\tau}\Big|_{z=0} &= \int_0^\infty \left[\hat{E}_{0v_1}(\lambda) \vec{V}_0^{(1)}(\lambda\rho) + \hat{E}_{0v_2}(\lambda) \vec{V}_0^{(2)}(\lambda\rho) \right] d\lambda, \\ \vec{H}_{0\tau}\Big|_{z=0} &= \int_0^\infty \left[\hat{H}_{0v_1}(\lambda) \vec{V}_0^{(1)}(\lambda\rho) + \hat{H}_{0v_2}(\lambda) \vec{V}_0^{(2)}(\lambda\rho) \right] d\lambda,\end{aligned}$$

где $\hat{E}'_{1v_1} = x_1(\lambda)$, $\hat{E}'_{1v_2} = \bar{v}_0(\lambda) y_1(\lambda)$, $\bar{v}_0(\lambda) = \frac{v_0(\lambda)}{k_0}$, $\hat{H}'_{1v_1} = h_0 y_1(\lambda)$, $\hat{H}'_{1v_2} = h_0 \bar{v}_0(\lambda) x_1(\lambda)$;

$$\hat{E}_{2v_1} = x_2(\lambda) F(\lambda), \quad \hat{E}_{2v_2} = -\bar{v}_0(\lambda) y_2(\lambda) F(\lambda), \quad F(\lambda) = e^{-v_0(\lambda)\Delta},\tag{12}$$

$$\hat{H}_{2v_1} = h_0 y_2(\lambda) F(\lambda), \quad \hat{H}_{2v_2} = -h_0 \bar{v}_0(\lambda) x_2(\lambda) F(\lambda);$$

$$\hat{E}_{0v_1} = a(\lambda), \quad \hat{E}_{0v_2} = -\bar{v}_0(\lambda) b(\lambda), \quad \hat{H}_{0v_1} = h_0 b(\lambda), \quad \hat{H}_{0v_2} = -h_0 \bar{v}_0(\lambda) a(\lambda);$$

$$\hat{E}_{1v_j} = \hat{E}_{0v_j} + \hat{E}'_{1v_j}, \quad \hat{H}_{1v_j} = \hat{H}_{0v_j} + \hat{H}'_{1v_j}.$$

3. Решение краевой задачи экранирования

Краевая задача (3)–(6) является трехобластной краевой задачей с классическими граничными условиями сопряжения (3) на плоскостях раздела сред Γ_1 и Γ_2 . В статье рассматриваются поля с осевой симметрией специального вида. Это позволяет сформулировать задачу (3)–(6) в виде двухобластной краевой задачи для областей D_1 и D_2 с нелокальными граничными условиями, связывающими поля по обе стороны экрана D [18].

Краевая задача 2. Требуется при заданном первичном поле \vec{E}_0, \vec{H}_0 определить поля $\vec{E}'_j, \vec{H}'_j \in C^1(D_1) \cap C(\bar{D}_1)$, $\vec{E}_2, \vec{H}_2 \in C^1(D_2) \cap C(\bar{D}_2)$, для которых выполнены уравнения

$$\text{rot } \vec{E}_j = i\omega\mu_0 \vec{H}_j, \quad \text{rot } \vec{H}_j = -i\omega\varepsilon_0 \vec{E}_j \quad \text{в } D_j,\tag{13}$$

нелокальные граничные условия

$$\vec{W}_1\Big|_{\Gamma_1} = \hat{A}^{-1} \vec{W}_2\Big|_{\Gamma_2}\tag{14}$$

и условия излучения на бесконечности (6).

Здесь

$$\hat{A}^{-1} = \hat{A}(-\Delta) = \begin{pmatrix} A_{11} & -A_{12} & A_{13} & -A_{14} \\ -A_{21} & A_{22} & -A_{23} & A_{24} \\ A_{31} & -A_{32} & A_{33} & -A_{34} \\ -A_{41} & A_{42} & -A_{43} & A_{44} \end{pmatrix}, \quad \vec{W}_j = \begin{pmatrix} \hat{E}_{jv_1} \\ \hat{E}_{jv_2} \\ \hat{H}_{jv_1} \\ \hat{H}_{jv_2} \end{pmatrix},\tag{15}$$

$$\begin{aligned}
 A_{11} &= p(p_1 C_2 - p_2 C_1), & A_{12} &= p \left(p_2 \frac{q_1}{l_1} S_1 - p_1 \frac{q_2}{l_2} S_2 \right), \\
 A_{13} &= Z_0 p (C_1 - C_2), & A_{14} &= Z_0 p \left(\frac{q_2}{l_2} S_2 - \frac{q_1}{l_1} S_1 \right), \\
 A_{21} &= p \left(p_2 \frac{l_1}{q_1} S_1 - p_1 \frac{l_2}{q_2} S_2 \right), & A_{22} &= p (p_1 C_2 - p_2 C_1), \\
 A_{23} &= Z_0 p \left(\frac{l_2}{q_2} S_2 - \frac{l_1}{q_1} S_1 \right), & A_{24} &= Z_0 p (C_1 - C_2), \\
 A_{31} &= \frac{pp_1 p_2}{Z_0} (C_2 - C_1), & A_{32} &= \frac{pp_1 p_2}{Z_0} \left(\frac{q_1}{l_1} S_1 - \frac{q_2}{l_2} S_2 \right), \\
 A_{33} &= p (p_1 C_1 - p_2 C_2), & A_{34} &= p \left(p_2 \frac{q_2}{l_2} S_2 - p_1 \frac{q_1}{l_1} S_1 \right), \\
 A_{41} &= \frac{pp_1 p_2}{Z_0} \left(\frac{l_1}{q_1} S_1 - \frac{l_2}{q_2} S_2 \right), & A_{42} &= \frac{pp_1 p_2}{Z_0} (C_2 - C_1), \\
 A_{43} &= p \left(p_2 \frac{l_2}{q_2} S_2 - p_1 \frac{l_1}{q_1} S_1 \right), & A_{44} &= p (p_1 C_1 - p_2 C_2); \\
 C_j &= \text{ch}(k_0 v_j \Delta), & S_j &= \text{sh}(k_0 v_j \Delta).
 \end{aligned} \tag{16}$$

Матричные коэффициенты (16) определяются через параметры уравнения (3):

$$\begin{aligned}
 \varepsilon &= \varepsilon_r \varepsilon_0, \quad \mu = \mu_r \mu_0, \quad G = G_r \sqrt{\varepsilon_0 \mu_0}, \quad Z = Z_r \sqrt{\varepsilon_0 \mu_0}, \quad l_j = \frac{v_j}{k_j}, \quad q_j = \frac{g}{k_j g_j}, \\
 g &= \varepsilon_r \mu_r - Z_r G_r, \quad g_j = f_j - \frac{1}{2} a, \quad a = i(G_r - Z_r), \quad f_j = (-1)^j f_0, \quad j=1, 2, \\
 f_0 &= \sqrt{\varepsilon_r \mu_r - \frac{1}{4} (G_r + Z_r)^2}, \quad 0 \leq \arg f_0 < \pi, \quad p_j = \frac{1}{\mu_r} \left(i \frac{g}{g_j} - Z_r \right), \\
 k_j &= \sqrt{g + \frac{1}{2} a^2 + a f_j}, \quad 0 \leq \arg k_j < \pi, \\
 v_j &= \sqrt{\bar{\lambda}^2 - k_j^2}, \quad -\frac{\pi}{2} \leq \arg v_j < \frac{\pi}{2}, \quad \bar{\lambda} = \frac{\lambda}{k_0}, \quad p = \frac{1}{p_1 - p_2}.
 \end{aligned}$$

Заметим, что граничные условия (14) сформированы для амплитуд (12) подынтегральных выражений (11). При этом задача экранирования (13), (14) эквивалентна краевой задаче (3)–(6). Видно, что для обратной матрицы (15) $\hat{A}^{-1} = \left\{ (-1)^{s+l} A_{sl} \right\}$.

Удовлетворим граничным условиям (14), подставив в них выражения (12). Получим систему алгебраических уравнений для определения функций $x_j(\lambda)$, $y_j(\lambda)$, входящих в интегралы (7), (8):

$$\begin{aligned}
& \left(A_{11}\hat{E}_{2v_1} - A_{12}\hat{E}_{2v_2} + A_{13}\hat{H}_{2v_1} - A_{14}\hat{H}_{2v_2} \right) = x_1 + a, \\
& \frac{1}{\bar{v}_0} \left(-A_{21}\hat{E}_{2v_1} + A_{22}\hat{E}_{2v_2} - A_{23}\hat{H}_{2v_1} + A_{24}\hat{H}_{2v_2} \right) = y_1 - b, \\
& iZ_0 \left(A_{31}\hat{E}_{2v_1} - A_{32}\hat{E}_{2v_2} + A_{33}\hat{H}_{2v_1} - A_{34}\hat{H}_{2v_2} \right) = y_1 + b, \\
& \frac{iZ_0}{\bar{v}_0} \left(-A_{41}\hat{E}_{2v_1} + A_{42}\hat{E}_{2v_2} - A_{43}\hat{H}_{2v_1} + A_{44}\hat{H}_{2v_2} \right) = x_1 - a.
\end{aligned} \tag{17}$$

Вычитая из первого равенства (17) четвертое, а из третьего равенства второе, приходим к двум уравнениям, не содержащим неизвестные x_1 и y_1 :

$$\begin{aligned}
& \left(A_{11} + \frac{iZ_0}{\bar{v}_0} A_{41} \right) \hat{E}_{2v_1} - \left(A_{12} + \frac{iZ_0}{\bar{v}_0} A_{42} \right) \hat{E}_{2v_2} + \\
& + \left(A_{13} + \frac{iZ_0}{\bar{v}_0} A_{43} \right) \hat{H}_{2v_1} - \left(A_{14} + \frac{iZ_0}{\bar{v}_0} A_{44} \right) \hat{H}_{2v_2} = 2a; \\
& \left(iZ_0 A_{31} + \frac{1}{\bar{v}_0} A_{21} \right) \hat{E}_{2v_1} - \left(iZ_0 A_{32} + \frac{1}{\bar{v}_0} A_{22} \right) \hat{E}_{2v_2} + \\
& + \left(iZ_0 A_{33} + \frac{1}{\bar{v}_0} A_{23} \right) \hat{H}_{2v_1} - \left(iZ_0 A_{34} + \frac{1}{\bar{v}_0} A_{24} \right) \hat{H}_{2v_2} = 2b.
\end{aligned} \tag{18}$$

Учтем в этих равенствах формулы (12) и определим неизвестные x_2, y_2 , разрешая систему уравнений (18). Получим

$$x_2(\lambda) = \frac{2G(\lambda)}{D(\lambda)} (a(\lambda)Q_{22}(\lambda) - b(\lambda)Q_{12}(\lambda)), \quad y_2(\lambda) = \frac{2G(\lambda)}{D(\lambda)} (b(\lambda)Q_{11}(\lambda) - a(\lambda)Q_{21}(\lambda)), \tag{19}$$

где

$$D(\lambda) = Q_{11}(\lambda)Q_{22}(\lambda) - Q_{12}(\lambda)Q_{21}(\lambda),$$

$$G = e^{v_0(\lambda)\Delta}, \quad Q_{11} = A_{11} + h_0\bar{v}_0A_{14} + \frac{iZ_0}{\bar{v}_0}A_{41} + A_{44}, \quad Q_{12} = \bar{v}_0A_{12} + h_0A_{13} + iZ_0A_{42} + \frac{1}{\bar{v}_0}A_{43},$$

$$Q_{21} = \frac{1}{\bar{v}_0}A_{21} + h_0A_{24} + iZ_0A_{31} + \bar{v}_0A_{34}, \quad Q_{22} = A_{22} + \frac{h_0}{\bar{v}_0}A_{23} + iZ_0\bar{v}_0A_{32} + A_{33}.$$

Из уравнений (17) определим коэффициенты отраженного поля (8):

$$\begin{aligned}
x_1 &= F \left[(A_{11} + h_0\bar{v}_0A_{14})x_2 + (\bar{v}_0A_{12} + h_0A_{13})y_2 \right] - a; \\
y_1 &= -\frac{F}{\bar{v}_0} \left[(A_{21} + h_0\bar{v}_0A_{24})x_2 + (\bar{v}_0A_{22} + h_0A_{23})y_2 \right] + b.
\end{aligned} \tag{20}$$

Используя формулы (19), (20), определим решение краевой задачи (13), (14):

$$\bar{E}_2 = \int_0^\infty [x_2(\lambda)J_1(\lambda\rho)\bar{e}_\varphi + y_2(\lambda)(\bar{v}_0(\lambda)J_1(\lambda\rho)\bar{e}_\rho + \bar{\lambda}J_0(\lambda\rho)\bar{e}_z)] e^{-v_0(\lambda)z} d\lambda,$$

$$H_2 = h_0 \int_0^{\infty} [y_2(\lambda)J_1(\lambda\rho)\vec{e}_\varphi + x_2(\lambda)(\bar{v}_0(\lambda)J_1(\lambda\rho)\vec{e}_\rho + \bar{\lambda}J_0(\lambda\rho)\vec{e}_z)]e^{-v_0(\lambda)z}d\lambda, \quad z > \Delta; \quad (21)$$

$$\vec{E}'_1 = \int_0^{\infty} [x_1(\lambda)J_1(\lambda\rho)\vec{e}_\varphi + y_1(\lambda)(-\bar{v}_0(\lambda)J_1(\lambda\rho)\vec{e}_\rho + \bar{\lambda}J_0(\lambda\rho)\vec{e}_z)]e^{v_0(\lambda)z}d\lambda,$$

$$\vec{H}'_1 = h_0 \int_0^{\infty} [y_1(\lambda)J_1(\lambda\rho)\vec{e}_\varphi + x_1(\lambda)(-\bar{v}_0(\lambda)J_1(\lambda\rho)\vec{e}_\rho + \bar{\lambda}J_0(\lambda\rho)\vec{e}_z)]e^{v_0(\lambda)z}d\lambda, \quad z < 0.$$

4. Алгоритм вычисления поля, проходящего через экран

Для численного исследования поля \vec{E}_2 , представленного в интегральном виде (21), воспользуемся следующим алгоритмом.

1. Исходные данные:

f – частота;

Δ – толщина экрана;

ε_r – относительная комплексная диэлектрическая проницаемость;

μ_r – относительная комплексная магнитная проницаемость;

G_r, Z_r – относительные комплексные параметры киральности;

c – скорость света;

A, B – произвольные комплексные постоянные;

α_1, α_2 – постоянные ($\alpha_1 > 0, \alpha_2 > 0$);

z, ρ – цилиндрические координаты точки $M(z > \Delta, 0 \leq \rho < \infty)$;

$\bar{\lambda}$ ($0 \leq \bar{\lambda} < \infty$) – переменная интегрирования;

n ($n = 0, 1, 2, \dots$) – параметр пучка волн;

h – расстояние до экрана.

2. Вспомогательные величины:

$$\omega = 2\pi f, \quad k_0 = \frac{\omega}{c}, \quad Z_0 = \sqrt{\mu_0/\varepsilon_0},$$

$$\bar{\Delta} = k_0\Delta, \quad \bar{h} = k_0h, \quad \bar{\rho} = k_0\rho, \quad \bar{z} = k_0z,$$

$$g = \varepsilon_r\mu_r - Z_rG_r, \quad a = i(G_r - Z_r),$$

$$f_0 = \sqrt{\varepsilon_r\mu_r - \frac{1}{4}(G_r + Z_r)^2}, \quad 0 \leq \arg f_0 < \pi, \quad f_j = (-1)^j f_0,$$

$$k_j = \sqrt{g + \frac{1}{2}a^2 + af_j}, \quad 0 \leq \arg k_j < \pi, \quad j = 1, 2,$$

$$g_j = f_j - \frac{1}{2}a, \quad q_j = \frac{g}{k_jg_j}, \quad p_j = \frac{1}{\mu_r} \left(i \frac{g}{g_j} - Z_r \right),$$

$$p = \frac{1}{p_1 - p_2}, \quad \bar{\alpha}_j = \alpha_j/k_0.$$

3. Вспомогательные функции переменной $\bar{\lambda}$ ($0 \leq \bar{\lambda} < \infty$):

$$v_j(\bar{\lambda}) = \sqrt{\bar{\lambda}^2 - k_j^2}, \quad -\frac{\pi}{2} \leq \arg v_j < \frac{\pi}{2}, \quad l_j(\bar{\lambda}) = v_j(\bar{\lambda})/k_j;$$

$$w_0(\bar{\lambda}) = \sqrt{|\bar{\lambda}^2 - 1|},$$

$$g_n(\bar{\lambda}) = (\bar{\lambda}^2 - 2n)L_n(\bar{\lambda}^2) + 2nL_{n-1}(\bar{\lambda}^2),$$

$$\tilde{C}_1(\bar{\lambda}) = \frac{1}{2}(1 + \exp(-2v_1(\bar{\lambda})\bar{\Delta})),$$

$$\tilde{C}_2(\bar{\lambda}) = \frac{1}{2}[\exp((v_2(\bar{\lambda}) - v_1(\bar{\lambda}))\bar{\Delta}) + \exp(-(v_2(\bar{\lambda}) + v_1(\bar{\lambda}))\bar{\Delta})],$$

$$\tilde{S}_1(\bar{\lambda}) = \frac{1}{2}(1 - \exp(-2v_1(\bar{\lambda})\bar{\Delta})),$$

$$\tilde{S}_2(\bar{\lambda}) = \frac{1}{2}[\exp((v_2(\bar{\lambda}) - v_1(\bar{\lambda}))\bar{\Delta}) - \exp(-(v_2(\bar{\lambda}) + v_1(\bar{\lambda}))\bar{\Delta})].$$

$\tilde{A}_{s,l}(\bar{\lambda})$ ($s, l = 1, 2, 3, 4$) – матричные функции, которые вычисляются по формулам (16) после замены функций $C_j(\bar{\lambda}), S_j(\bar{\lambda})$ на функции $\tilde{C}_j(\bar{\lambda}), \tilde{S}_j(\bar{\lambda})$.

4. Вспомогательные функции на отрезках $0 \leq \bar{\lambda} < 1, 1 < \bar{\lambda} < \infty$:

$$A_0^{(1)}(\bar{\lambda}) = k_0 w_0(\bar{\lambda}) A g_n \left(\frac{\bar{\lambda}}{\alpha_1} \right) \exp \left(i w_0(\bar{\lambda}) \bar{h} - \frac{1}{2} \left(\frac{\bar{\lambda}}{\alpha_1} \right)^2 \right), \quad 0 \leq \bar{\lambda} < 1,$$

$$A_0^{(2)}(\bar{\lambda}) = k_0 w_0(\bar{\lambda}) A g_n \left(\frac{\bar{\lambda}}{\alpha_1} \right) \exp \left(-w_0(\bar{\lambda}) \bar{h} - \frac{1}{2} \left(\frac{\bar{\lambda}}{\alpha_1} \right)^2 \right), \quad 1 < \bar{\lambda} < \infty;$$

$$B_0^{(1)}(\bar{\lambda}) = i k_0 B g_n \left(\frac{\bar{\lambda}}{\alpha_2} \right) \exp \left(i w_0(\bar{\lambda}) \bar{h} - \frac{1}{2} \left(\frac{\bar{\lambda}}{\alpha_2} \right)^2 \right), \quad 0 \leq \bar{\lambda} < 1,$$

$$B_0^{(2)}(\bar{\lambda}) = k_0 B g_n \left(\frac{\bar{\lambda}}{\alpha_2} \right) \exp \left(-w_0(\bar{\lambda}) \bar{h} - \frac{1}{2} \left(\frac{\bar{\lambda}}{\alpha_2} \right)^2 \right), \quad 1 \leq \bar{\lambda} < \infty;$$

$$G^{(1)}(\bar{\lambda}) = \exp(-i w_0(\bar{\lambda}) + v_1(\bar{\lambda})\bar{\Delta}), \quad 0 \leq \bar{\lambda} < 1,$$

$$G^{(2)}(\bar{\lambda}) = \exp((w_0(\bar{\lambda}) - v_1(\bar{\lambda}))\bar{\Delta}), \quad 1 \leq \bar{\lambda} < \infty;$$

$$Q_{11}^{(1)}(\bar{\lambda}) = w_0(\bar{\lambda}) \left(\tilde{C}_1(\bar{\lambda}) + \tilde{C}_2(\bar{\lambda}) - \frac{w_0(\bar{\lambda})}{Z_0} \tilde{A}_{14}(\bar{\lambda}) \right) - Z_0 A_{41}(\bar{\lambda}),$$

$$Q_{12}^{(1)}(\bar{\lambda}) = i w_0(\bar{\lambda}) (p(1 + p_1 p_2) (\tilde{C}_2(\bar{\lambda}) - \tilde{C}_1(\bar{\lambda})) - w_0(\bar{\lambda}) \tilde{A}_{12}(\bar{\lambda})) + i \bar{A}_{43}(\bar{\lambda}),$$

$$Q_{21}^{(1)}(\bar{\lambda}) = i w_0(\bar{\lambda}) (p(1 + p_1 p_2) (\tilde{C}_2(\bar{\lambda}) - \tilde{C}_1(\bar{\lambda})) - w_0(\bar{\lambda}) \tilde{A}_{34}(\bar{\lambda})) + i \bar{A}_{21}(\bar{\lambda}),$$

$$\begin{aligned}
 Q_{22}^{(1)}(\bar{\lambda}) &= w_0(\bar{\lambda})(\tilde{C}_1(\bar{\lambda}) + \tilde{C}_2(\bar{\lambda}) + Z_0 w_0(\bar{\lambda}) \tilde{A}_{32}(\bar{\lambda})) + \frac{1}{Z_0} \tilde{A}_{23}(\bar{\lambda}), \quad 0 \leq \bar{\lambda} < 1; \\
 Q_{11}^{(2)}(\bar{\lambda}) &= w_0(\bar{\lambda}) \left(\tilde{C}_1(\bar{\lambda}) + \tilde{C}_2(\bar{\lambda}) - \frac{i w_0(\bar{\lambda})}{Z_0} \tilde{A}_{14}(\bar{\lambda}) \right) + i Z_0 A_{41}(\bar{\lambda}), \\
 Q_{12}^{(2)}(\bar{\lambda}) &= w_0(\bar{\lambda}) (i p (1 + p_1 p_2) (\tilde{C}_2(\bar{\lambda}) - \tilde{C}_1(\bar{\lambda})) + w_0(\bar{\lambda}) \tilde{A}_{12}(\bar{\lambda})) + \bar{A}_{43}(\bar{\lambda}), \\
 Q_{21}^{(2)}(\bar{\lambda}) &= w_0(\bar{\lambda}) (i p (1 + p_1 p_2) (\tilde{C}_2(\bar{\lambda}) - \tilde{C}_1(\bar{\lambda})) + w_0(\bar{\lambda}) \tilde{A}_{34}(\bar{\lambda})) + \bar{A}_{21}(\bar{\lambda}), \\
 Q_{22}^{(2)}(\bar{\lambda}) &= w_0(\bar{\lambda}) (\tilde{C}_1(\bar{\lambda}) + \tilde{C}_2(\bar{\lambda}) + i Z_0 w_0(\bar{\lambda}) \tilde{A}_{32}(\bar{\lambda})) - \frac{i}{Z_0} \tilde{A}_{23}(\bar{\lambda}), \quad 1 \leq \bar{\lambda} < \infty; \\
 Q^{(1)}(\bar{\lambda}) &= 2G^{(1)}(\bar{\lambda}) / [\tilde{Q}_{11}(\bar{\lambda}) \tilde{Q}_{22}(\bar{\lambda}) - \tilde{Q}_{12}(\bar{\lambda}) \tilde{Q}_{21}(\bar{\lambda})], \quad 0 \leq \bar{\lambda} < 1, \\
 Q^{(2)}(\bar{\lambda}) &= 2G^{(2)}(\bar{\lambda}) / [\tilde{Q}_{11}(\bar{\lambda}) \tilde{Q}_{22}(\bar{\lambda}) - \tilde{Q}_{12}(\bar{\lambda}) \tilde{Q}_{21}(\bar{\lambda})], \quad 1 \leq \bar{\lambda} < \infty; \\
 x_2^{(1)}(\bar{\lambda}) &= Q^{(1)}(\bar{\lambda}) (A_0^{(1)}(\bar{\lambda}) \tilde{Q}_{22}^{(1)}(\bar{\lambda}) - B_0^{(1)}(\bar{\lambda}) \tilde{Q}_{12}^{(1)}(\bar{\lambda})), \quad 0 \leq \bar{\lambda} < 1, \\
 y_2^{(1)}(\bar{\lambda}) &= Q^{(1)}(\bar{\lambda}) (B_0^{(1)}(\bar{\lambda}) \tilde{Q}_{11}^{(1)}(\bar{\lambda}) - A_0^{(1)}(\bar{\lambda}) \tilde{Q}_{21}^{(1)}(\bar{\lambda})), \quad 0 \leq \bar{\lambda} < 1; \\
 x_2^{(2)}(\bar{\lambda}) &= Q^{(2)}(\bar{\lambda}) (A_0^{(2)}(\bar{\lambda}) \tilde{Q}_{22}^{(2)}(\bar{\lambda}) - B_0^{(2)}(\bar{\lambda}) \tilde{Q}_{12}^{(2)}(\bar{\lambda})), \quad 1 < \bar{\lambda} < \infty, \\
 y_2^{(2)}(\bar{\lambda}) &= Q^{(2)}(\bar{\lambda}) (B_0^{(2)}(\bar{\lambda}) \tilde{Q}_{11}^{(2)}(\bar{\lambda}) - A_0^{(2)}(\bar{\lambda}) \tilde{Q}_{21}^{(2)}(\bar{\lambda})), \quad 1 < \bar{\lambda} < \infty.
 \end{aligned}$$

5. Вычисление компонент электрического поля $\vec{E}_2 = E_{2\phi} \vec{e}_\phi + E_{2\rho} \vec{e}_\rho + E_{2z} \vec{e}_z$:

$$\begin{aligned}
 E_{2\phi}(\rho, z) &= \int_0^1 x_2^{(1)}(\bar{\lambda}) J_1(\bar{\lambda} \bar{\rho}) e^{i w_0(\bar{\lambda}) z} d\bar{\lambda} + \int_1^\infty x_2^{(2)}(\bar{\lambda}) J_1(\bar{\lambda} \bar{\rho}) e^{-w_0(\bar{\lambda}) z} d\bar{\lambda}, \\
 E_{2\rho}(\rho, z) &= -i \int_0^1 y_2^{(1)}(\bar{\lambda}) w_0(\bar{\lambda}) J_1(\bar{\lambda} \bar{\rho}) e^{i w_0(\bar{\lambda}) z} d\bar{\lambda} + \int_1^\infty y_2^{(2)}(\bar{\lambda}) w_0(\bar{\lambda}) J_1(\bar{\lambda} \bar{\rho}) e^{-w_0(\bar{\lambda}) z} d\bar{\lambda}, \quad (22) \\
 E_{2z}(\rho, z) &= \int_0^1 y_2^{(1)}(\bar{\lambda}) \bar{\lambda} J_0(\bar{\lambda} \bar{\rho}) e^{i w_0(\bar{\lambda}) z} d\bar{\lambda} + \int_1^\infty y_2^{(2)}(\bar{\lambda}) \bar{\lambda} J_0(\bar{\lambda} \bar{\rho}) e^{-w_0(\bar{\lambda}) z} d\bar{\lambda}.
 \end{aligned}$$

5. Вычислительный эксперимент

Для расчетов рассмотрен пучок Лагерра – Гаусса \vec{E}_0 на частоте $f = 3 \cdot 10^8$ Гц, ортогонально падающий на экран $D(0 < z < \Delta)$, $\Delta = 0,2$ м, и распространяющийся в положительном направлении оси $0z$. За начальное сечение пучка выбрана плоскость $z = -h$ ($h = 0,1$ м), в которой компоненты поля представляются аналитически в достаточно простом виде (2).

На рис. 1 модули компонент электрического поля пучка (1) изображены в четырех ортогональных сечениях ($z = \text{const}$), когда экран D отсутствует. Выбраны следующие параметры пучка: $f = 3 \cdot 10^8$ Гц, $h = 0,1$ м, $A = 1$, $B = 0$, $\alpha_1 = 2 \text{ м}^{-1}$, $\alpha_2 = 2 \text{ м}^{-1}$, $n = 1$, $n = 5$. В этом случае вектор \vec{E}_0 параллелен экрану D (TE -поляризация), компоненты $E_{0\rho}$, E_{0z} отсутствуют. Графики на рисунках являются образующими вращения с осью симметрии $0z$.

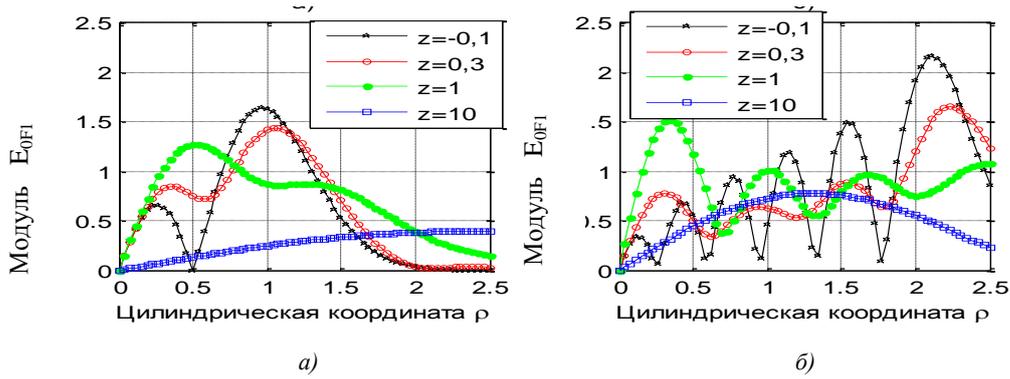


Рис. 1. Графики модуля компоненты $E_{0\varphi}$ электрического поля \vec{E}_0 при отсутствии экрана: а) $n=1$; б) $n=5$

На рис. 2 при аналогичных значениях параметров изображено изменение реального электрического поля $\text{Re}(\vec{E}_0 e^{-i\omega t})$ пучка в течение периода колебаний электромагнитного поля, $0 \leq \bar{t} \leq 1$, $\bar{t} = t/T$ – безразмерное время, $T = 1/f$ – период поля. Параметры пучка: $f = 3 \cdot 10^8$ Гц, $h = 0,1$ м, $A = 1$, $B = 0$, $\alpha_1 = 2 \text{ м}^{-1}$, $\alpha_2 = 2 \text{ м}^{-1}$, $n = 1$.

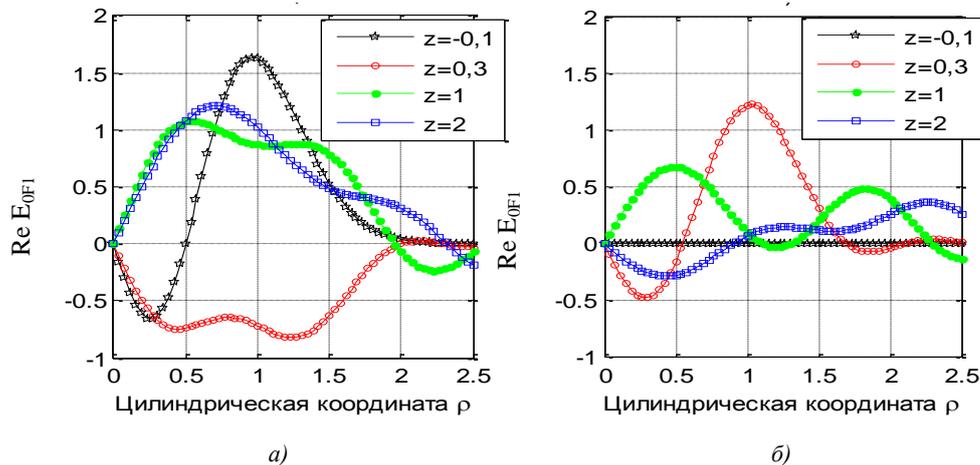


Рис. 2. Изменение структуры пучка волн \vec{E}_0 ($A = 1, B = 0, \alpha_1 = 2, n = 1$) в течение периода при отсутствии экрана: а) $\bar{t} = 0$; б) $\bar{t} = 0,25$

На рис. 3 и 4 представлены графики, описывающие структуру пучка волн (7) в полупространстве $D_2(z > \Delta)$ при прохождении пучка (1) через биизотропный экран D с различными значениями параметров киральности. На рис. 3 изображена структура пучка при наличии экрана из кирального метаматериала с параметрами $\Delta = 0,2$ м, $G_r = 3i$, $Z_r = 3i$, $\epsilon_r = -4$, $\mu_r = 2$. Рассмотрен пучок (1) при $f = 3 \cdot 10^8$ Гц, $h = 0,1$ м, $A = 1$, $B = 0$, $\alpha_1 = 2 \text{ м}^{-1}$, $n = 1$, представляющий собой TE -поляризованное поле. Заметим, что пучок (1) при прохождении через экран D преобразуется в поле \vec{E}_2 с компонентами (22), состоящее из двух поляризаций: $\vec{E}_2 = \vec{E}_2^{TE} + \vec{E}_2^{TH}$. На рис. 3, а показано TE -поляризованное поле $\vec{E}_2^{TE} = E_{2\varphi} \vec{e}_\varphi$, на рис. 3, б и в – компоненты TH -поляризованного поля $\vec{E}_2^{TH} = E_{2\rho} \vec{e}_\rho + E_{2z} \vec{e}_z$ (вектор \vec{H}_2^{TH} параллелен экрану D), а на рис. 3, г – модуль поля $|\vec{E}_2|$.

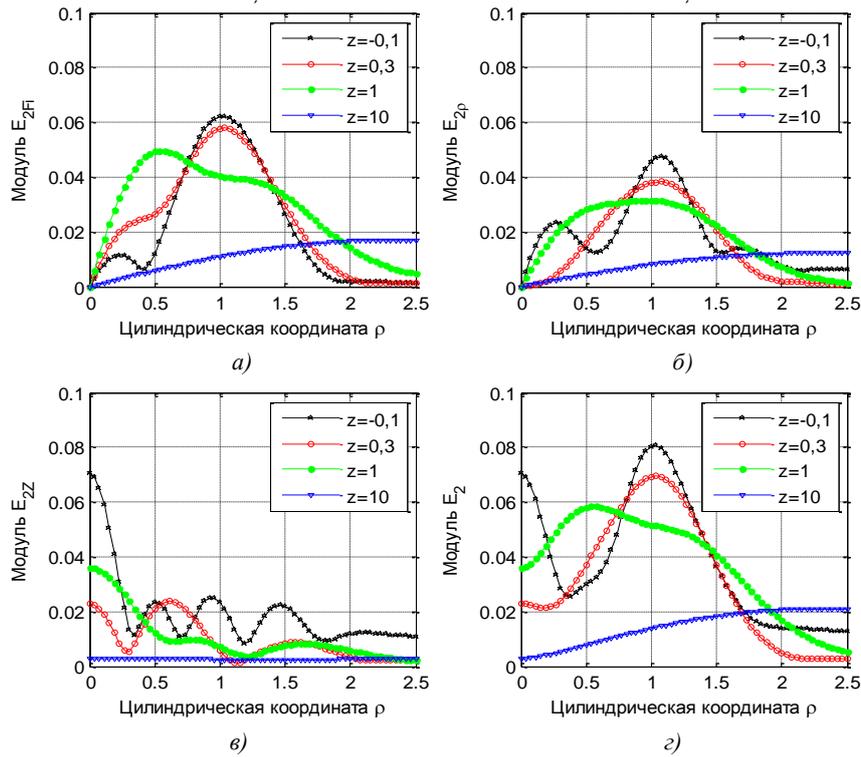


Рис. 3. Графики модуля компонент поля \vec{E}_2 ($A = 1, B = 0, \alpha_1 = 2, n = 1$) при наличии экрана из кирального метаматериала: а) $|E_{z\phi}|$; б) $|E_{z\rho}|$; в) $|E_{zz}|$; г) $|\vec{E}_2|$

На рис. 4 показаны результаты решения задачи (3)–(6) для параметров $f = 3 \cdot 10^9$, $\Delta = 0,05$ м, $G_r = 0,5i$, $Z_r = -0,5i$, $\epsilon_r = 4$, $\mu_r = 1$, $A = 1, B = 0, h = 0,1$ м, $\alpha_1 = 2 \text{ м}^{-1}$, $n = 1$.

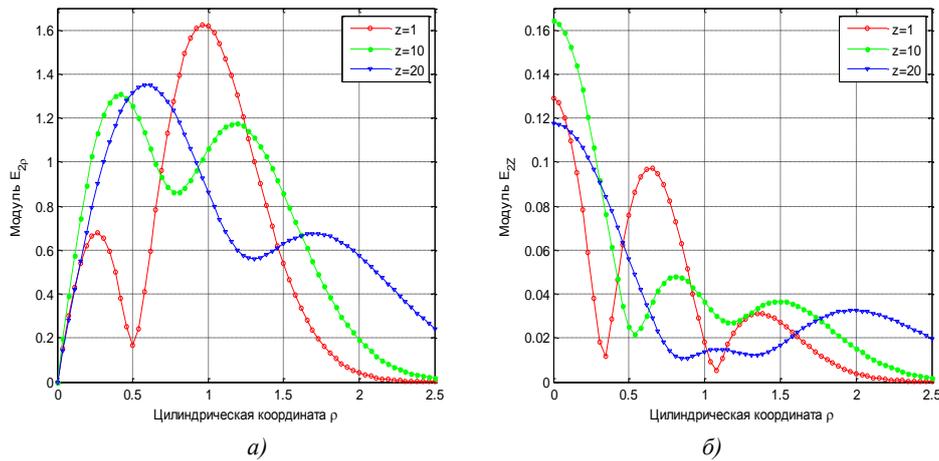


Рис. 4. Модули компонент поля \vec{E}_2 при наличии экрана из кирального материала: а) $|E_{z\rho}|$; б) $|E_{zz}|$

В работе [19] рассмотрен экран с аналогичными параметрами, через который ортогонально экрану проходит плоская TE -поляризованная волна. Показано, что в результате прохождения TE -поляризованная плоская волна преобразуется в TH -поляризованную плоскую волну. Аналогичная ситуация наблюдается и для пучка волн при ортогональном падении на экран: TE -поляризованный пучок волн (первичное поле) при прохождении через экран преобразуется в TH -поляризованный пучок. Компонента $E_{z\phi}$ (TE -поле) в полупространстве D_2 практически

равна нулю (в D_1 для первичного поля $E_{0p} \neq 0$, $E_{0r} \equiv 0$, $E_{0z} \equiv 0$). На рис. 4 изображены компоненты TH -поляризованного пучка волн в D_2 . При этом преобладает компонента E_{2p} . Сравнение рисунков 3, а и 4, а позволяет заключить, что при увеличении частоты поля пучок волн в направлении оси Oz расплывается гораздо медленнее.

Численно исследовано решение краевой задачи (3)–(6) для параметров $f = 3 \cdot 10^8$ Гц, $h = 0,1$ м, $A = 1$, $B = 0$, $\alpha_1 = 2$ м⁻¹, $n = 1$, $\Delta = 0,2$ м, $G_r = 3i$, $Z_r = 3i$, $\varepsilon_r = -4$, $\mu_r = 2$ с фокусирующим параметром $\tau = 0$ [9]. Показано, что при прохождении пучка (1) через экран D пучок сдвигается вдоль оси Oz на расстояние 0,4 м и полностью сохраняет вид. Начало пучка в сечении $z = -0,1$ переходит в сечение $z = 0,3$.

Отметим, что в публикации [20] аналитически исследованы пучки волн с зависимостью от азимутального угла φ .

Заключение

В работе сформулирована краевая задача прохождения кругового пучка электромагнитных волн через плоский экран из биизотропного материала с использованием нелокальных двухсторонних граничных условий специального типа, связывающих поля по обе стороны экрана. Изучены пучки волн Лагерра – Гаусса, которые являются произвольными комбинациями TE - и TH -поляризованных пучков, распространяющихся ортогонально плоскому экрану с произвольными комплексными диэлектрической и магнитной проницаемостями и параметрами киральности. Методика решения задачи основана на представлении первичного и результирующего пучков волн через базисные цилиндрические электромагнитные поля. В результате компоненты поля, прошедшего через экран, представлены в виде несобственных интегралов, содержащих функции Бесселя.

Разработана программа и численно исследованы свойства экранов из специальных киральных материалов. Показано, что с помощью кирального экрана TE -поляризованные пучки волн преобразуются в TH -поляризованные пучки и наоборот. С помощью экрана из кирального метаматериала пучок волн сдвигается вдоль оси пучка с полным сохранением структуры.

Работа выполнена по заданию № 20110838 ГПНИ «Информатика и космос».

Список литературы

1. Виноградов, А.П. Электродинамика композитных материалов / А.П. Виноградов. – М. : Эдиториал УРСС, 2001. – 206 с.
2. Electromagnetic Waves in Chiral and Bi-isotropic Media / I.V. Lindell [et al.]. – Boston : Artech House, 1994. – 324 p.
3. Иванов, О.В. Распространение электромагнитных волн в анизотропных и бианизотропных слоистых структурах / О.В. Иванов. – Ульяновск : УлГТУ, 2010. – 262 с.
4. Гуляев, Ю.В. Метаматериалы: фундаментальные исследования и перспективы применения / Ю.В. Гуляев, А.Н. Лагарьков, С.А. Никитов // Вестник РАН. – 2008. – Т. 78, № 5. – С. 438–457.
5. Веселаго, В.Г. Электродинамика веществ с одновременно отрицательными значениями ε и μ / В.Г. Веселаго // Успехи физических наук. – 1967. – Т. 92, № 3. – С. 517.
6. Миттра, Р. Критический взгляд на метаматериалы / Р. Миттра // Радиотехника и электроника. – 2007. – Т. 52, № 9. – С. 1051–1058.
7. Ерофеев, В.Т. Алгоритм численного исследования экранирующих свойств многослойных экранов из композитных материалов / В.Т. Ерофеев, С.В. Малый // Информатика. – 2010. – № 4 (28). – С. 96–103.
8. Иванов, О.В. Распространение света в киральных слоистых средах. Метод матриц 4×4 / О.В. Иванов, Д.И. Семенов // Кристаллография. – 2000. – Т. 45, № 3. – С. 534–540.

9. 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.
10. Кухарчик, П.Д. Полное внутреннее отражение гауссова светового пучка / П.Д. Кухарчик, В.М. Сердюк, И.А. Титовицкий // ЖТФ. – 1999. – Т. 69, № 4. – С. 74–79.
11. Казак, Л.А. Формирование, суперпозиция и устойчивость вихревых оптических пучков различных порядков / Л.А. Казак, А.Л. Толстик // Вестник БГУ. Сер. 1. – 2010. – № 2. – С. 3–7.
12. Романов, О.Г. Самофокусировка и модуляционная неустойчивость оптических вихрей в резонансных средах / О.Г. Романов, А.Л. Толстик // Вестник БГУ. Сер. 1. – 2010. – № 3. – С. 13–17.
13. Шевченко, В.В. Геометрооптическая теория плоской линзы из кирального метаматериала / В.В. Шевченко // Радиотехника и электроника. – 2009. – Т. 54, № 6. – С. 696–700.
14. Иванов, О.В. Трансформация гауссова светового пучка при отражении вблизи угла Брюстера / О.В. Иванов, Д.И. Семенов // Оптика и спектроскопия. – 2002. – Т. 92, № 3. – С. 462–467.
15. Панов, В.П. О распространении волновых пучков в средах с изотропной комплексной диэлектрической проницаемостью / В.П. Панов, В.В. Приходько // Радиотехника и электроника. – 2007. – Т. 52, № 6. – С. 662–670.
16. Гончаренко, А.М. Оптические гауссовы пучки и солитоны / А.М. Гончаренко. – Минск : Беларуская навука, 2011. – 126 с.
17. Ерофеенко, В.Т. Аналитическое моделирование в электродинамике / В.Т. Ерофеенко, И.С. Козловская. – Минск : БГУ, 2010. – 304 с.
18. 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.
19. Ерофеенко, В.Т. Дифракция плоской электромагнитной волны на плоском слое из биизотропного материала / В.Т. Ерофеенко, С.В. Малый // Вестник БГУ. Сер.1. – 2010. – № 2. – С. 11–16.
20. Ерофеенко, В.Т. Краевая задача дифракции пучков электромагнитных волн на плоском экране из биизотропных материалов / В.Т. Ерофеенко // Известия НАН Беларуси. Сер. физ.-мат. наук. – 2012. – № 4. – С. 72–79.

Поступила 22.11.2012

¹ Учреждение БГУ «НИИ прикладных проблем математики и информатики»,
Минск, пр. Независимости, 4
e-mail: bsu_erofeenko@tut.by

² Высший государственный колледж связи,
Минск, ул. Ф. Скорины, 8, корп. 2
e-mail: valbandarenka@yandex.ru

V.T. Erofeenko, V.F. Bondarenko

TRANSFORMATION OF BEAMS OF ELECTROMAGNETIC WAVES PASSING THROUGH A CHIRAL METAMATERIAL SCREEN

The boundary value problem of penetration of beams of monochromatic electromagnetic fields through a flat screen of biisotropic material with arbitrary complex parameters of media is solved. The solution is represented in the integral form. The results of computational experiment for specific screens of chiral metamaterials are discussed.

УДК 004.048:519.816

М.К. Буза, В.С. Горшунов

СИСТЕМА МОДЕЛИРОВАНИЯ ПОВЕДЕНИЯ КЛИЕНТОВ ОРГАНИЗАЦИЙ

Исследуются система моделирования поведения клиентов, функционирование и архитектурные решения подобных систем. Предлагается механизм идентификации клиентов, основанный в отличие от известных на заданной целевой функции с использованием контрольных и динамических параметров.

Введение

С каждым годом бизнес становится все более конкурентным, идет объективный процесс его укрупнения. Становится понятно, что бизнес работает для клиента и поэтому должен быть клиентоориентированным. Это включает в себя привлечение новых клиентов, удержание старых и прогнозирование их оттока. Компании нужны ответы на вопросы, какие клиенты способны принести ей наибольшую прибыль, а какие отличаются наибольшей склонностью к смене компании как партнера. Не нарушая общности рассуждений, рассмотрим эти вопросы на примере банковской системы. Основными проблемами банка являются выделение сегмента клиентов, которые с высокой степенью вероятности могут отказаться от услуг банка, установление причин этого отказа и разработка выгодного с точки зрения банка и клиента решения, которое позволит их удержать.

Изучение поведения клиента показывает, что он становится клиентом банка, если банк может помочь решить его проблемы: получить кредит, приобрести мобильный интернет, отправиться в заграничную поездку, поддержать здоровье, получить образование и т. д. Изучать потребности клиента можно различными методами. Наиболее распространенными из них являются сбор и анализ семантических данных, в результате которых информация о всех клиентах обрабатывается и сегментируется. Именно сегментация клиентской базы позволяет увидеть реальную картину текущего состояния клиентов. Однако статистика дает только срез состояния на текущий период, а банку необходимо определить, что будет с клиентами завтра, через год или несколько лет. Десятилетний период кажется дальним горизонтом, поэтому многие компании о нем даже не задумываются. Однако лидеры рынка приходят к выводу о необходимости формирования клиентской базы на будущее. Ниже предлагается один из механизмов идентификации клиентов, основанный на моделировании их поведения [1–3].

1. Модель сегментирования клиентов

Цель модели – ранжирование множества клиентов на основе предложения пакета услуг с максимальной доходностью для банка. Сегментирование предложений банка зависит от принципа взаимодействия с клиентом. Например, клиент сам пришел в банк (reactive) или банк сделает ему предложение. Другим примером сегментирования может служить источник информирования клиента о предложении, которым может быть, например, партнер банка, либо рекламная sms-рассылка, либо сайт банка и т. д. [4].

Предположим, что банк разработал два пакета предложений:

1. Увеличение кредитного лимита клиента (Increase Credit Limit):
 - а) по инициативе клиента (Reactive: Increase Credit Limit);
 - б) по инициативе банка (Proactive: Increase Credit Limit).
2. Уменьшение годовой процентной ставки по кредиту (Reduce Annual Percentage Rate – Reduce APR):
 - а) по инициативе клиента (Reactive: Reduce APR);
 - б) по инициативе банка (Proactive: Reduce APR).

В модели предложения банка присутствуют как постоянные (контрольные) параметры, так и динамические.

Постоянные параметры включают:

1. *Пороговую благосклонность* клиента (Propensity Threshold) – минимальную возможную благосклонность клиента, при которой банк будет считать целесообразным делать данное предложение. Значения этого параметра изменяются в интервале от 0 до 1. В предложенной модели значения определены следующим образом:

0 – клиент не склонен рассматривать предложение банка;

0,25 – скорее нет, чем да;

0,5 – клиент размышляет над предложением;

0,75 – скорее да, чем нет;

1 – клиент готов принять предложение банка.

2. *Маркетинговый вес* (Marketing Weight) – показывает ценность данного предложения в стратегическом плане для банка. В рассматриваемой модели параметр принимает дискретные значения от 0 до 1 с шагом 0,25. Семантика значений предложений следующая:

0 – наименьшей ценности;

0,25 – малой ценности;

0,5 – средней ценности;

0,75 – высокой ценности;

1 – наибольшей ценности.

3. *Уровень сегментации* (Champion Challenger) – определяет пороговое значение (Champion), при превышении которого значением-кандидатом (Challenger) предложение может быть выставлено клиенту. Значения параметра изменяются в интервале от 0 до 1.

4. *Стимулирование клиента* (Incentive) – показывает, будет ли выполнено стимулирование клиента в контексте текущего предложения. Принимает значения:

0 – банк не будет стимулировать клиента;

1 – банк будет стимулировать клиента.

5. *Время действия предложения* (Valid To) – предельное время, до которого будет действовать предложение банка.

Динамические переменные параметры модели включают:

1. *Склонность клиента* принять предложение банка (Propensity) – интервал изменения от 0 до 1. Принимает значения, аналогичные значениям параметра пороговой склонности клиента.

2. *Доход* (Revenue) – вычисляется как сумма кредита, умноженная на процент кредита.

3. *Применимость предложения* (Fulfilled) – показывает, было ли принято решение о выдаче предложения банка клиенту. Принимает значения: 0 – не применимо, 1 – применимо.

Проанализируем структуру данных, состоящую из контрольных и динамических параметров модели предложения. Применительно к данному клиенту будем рассматривать два различных варианта этой структуры:

– *текущее состояние* (Champion), при котором для клиента рассматривается возможность выдачи предложения;

– *последующее состояние* (Challenger), при котором клиент принял предложение.

Принципы заполнения данных структур: контрольные параметры модели задаются маркетологами банка (пользователями модели); переменные параметры модели выбираются из базы накопленных данных, учитывающих предысторию работы с клиентами.

Из всех предложений, предлагаемых клиенту, выбирается то, у которого расчетное значение *веса коэффициента* (weighting factor – wf) вычисляется по формуле

$$wf = \begin{cases} 0 & \text{if propensity} < \{\text{Propensity Threshold}\}; \\ x_i & \text{if propensity} \geq \{\text{Propensity Threshold}\}, \end{cases}$$

где $x_i = \text{propensity} * \{\text{Marketing Weight}\}$.

В зависимости от весового коэффициента выбираем предложение, попадающее в интервал его значений:

```

if 0 <= wf <= 0,25 then Churn.Reactive.Reactive: Increase Credit Limit;
if 0,25 < wf <= 0,5 then Churn.Reactive.Reactive: Reduce APR;
if 0,5 < wf = 0,75 then Churn.Proactive.Proactive: Increase Credit Limit;
if 0,75 < wf <= 1 then Churn.Proactive.Proactive: Reduce APR.

```

После этого происходит обработка полученных результатов. Структуру, содержащую приведенные данные, назовем фактом (Fact).

В обработке участвуют следующие данные:

- 1) индекс предложения (Proposition Id) – уникально идентифицирует предложение, например Proactive: Increase Credit Limit;
- 2) источник информации клиента о предложении (Channel);
- 3) параметр доходности (Revenue) – берется из соответствующего поля модели предложения;
- 4) параметр применимости (Fulfilled) – берется из соответствующего поля модели предложения;
- 5) дата выполнения предложения.

Предложенная модель чувствительна к количеству используемых параметров: с увеличением числа параметров время моделирования увеличивается.

Целесообразно представлять модель на подмножестве языка XML [5, 6]. Это подмножество поддерживает работу как с примитивными типами (число, строка, булево значение, дата и т. д.), так и с массивами этих примитивных типов. Модель включает правила, которые связывают атрибуты клиентов (входные параметры). На выходе ее получаем реальные предложения с вероятностью принятия клиентом предложения банка.

2. Симуляция

Рассмотрим схему работы модели (рис. 1). На вход модели поступает информация о клиенте, значения контрольных параметров (Control Parameters values) и список предложений (Proposal list), на выходе из модели получаем предложение для клиента, которое в случае его принятия записывается в базу накопленной статистики как факт.

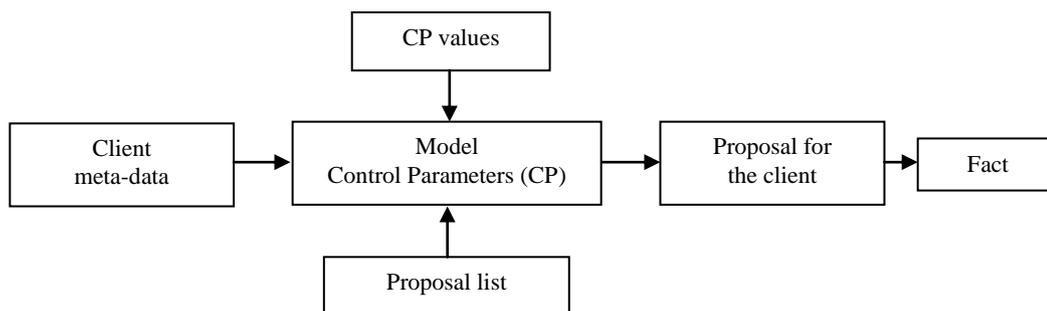


Рис. 1. Схема работы модели

На основании работы модели конкретному клиенту подбирается и делается конкретное предложение, которое он может принять или отклонить. Данные о каждом клиенте попадают в базу накопленной статистики и ранжируются по приведенным выше параметрам. Таким образом, формируется набор исторически накопленных фактов, который позволяет перейти к этапу прогнозирования доходности банка в зависимости от клиентуры банка и предложений клиентам. Этот процесс назовем процессом симуляции (Simulation). Он выполняется на основе исторически накопленных фактов с помощью генетических алгоритмов, которые строят предикативные значения показателей.

Перед началом работы модели произведем генерацию исторически накопленных фактов. Процедура генерации фактов называется *разогревом системы* (Warming up). Вместо генерации можно использовать реальные данные. Предположим, что имеется 10 000 исторически накопленных фактов.

Рассмотрим пример симуляции нашей модели. До начала симуляции модель имеет контрольные параметры (рис 2), которые характеризуют текущее поведение модели. Симулированные данные на их основе будем называть целевыми (target).

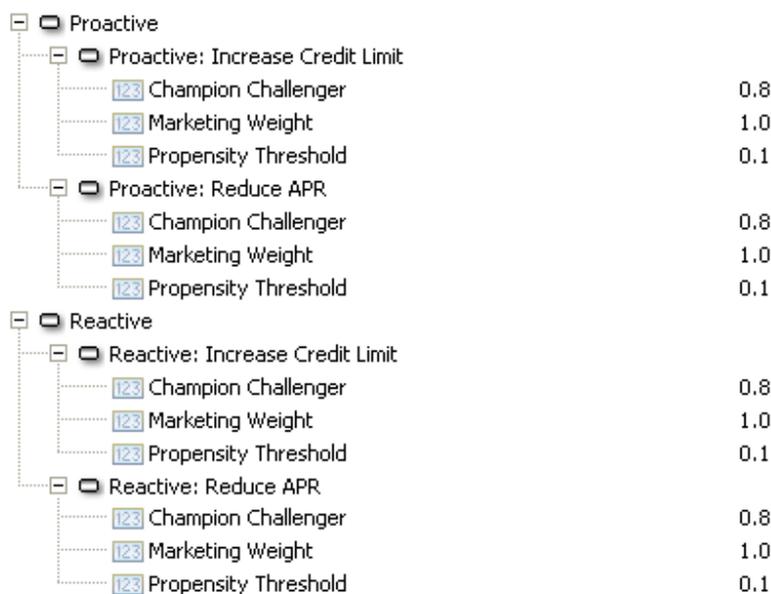


Рис. 2. Начальные значения контрольных параметров модели

Изменим маркетинговый вес для двух контрольных параметров (табл. 1).

Таблица 1

Изменения в маркетинговом весе проактивных предложений		
Предложение	Начальное значение	Новое значение
Proactive: Increase Credit Limit	0,0	0,25
Proactive: Reduce APR	0,0	0,5

Выполним процесс симуляции. Полученные в результате этого процесса данные будем называть симулированными (simulated). По его завершении строится таблица изменений в объеме принятых предложений (симулированных результатов) по сравнению с целевыми результатами (табл. 2).

Таблица 2

Изменения в объеме принятых предложений			
Предложение	Consumer: New	Consumer: Mature	Consumer: Developing
Reactive: Reduce APR	0	0	0
Reactive: Increase credit Limit	0	0	0
Proactive: Reduce APR	340	400	390
Proactive: Increase Credit Limit	210	140	70

Из табл. 2 виден рост принятых предложений клиентами при условии изменения контрольных параметров по сравнению с начальными параметрами модели. Данные в таблице по-

казывают изменения в количестве принятых предложений клиентами при условии изменения маркетингового веса по сравнению с прогнозируемым уровнем принятия предложений.

Проведем еще одну симуляцию, но уже с изменением маркетингового веса реактивного предложения (Reactive: Reduce APR) с начального значения 0,0 на новое значение, равное 0,75.

В результате получим следующие изменения в объеме принятых предложений по сравнению с прогнозируемым (табл. 3).

Таблица 3

Изменения в объеме принятых предложений			
Предложение	Consumer: New	Consumer: Mature	Consumer: Developing
Reactive: Reduce APR	40	120	60
Reactive: Increase credit Limit	60	10	90
Proactive: Reduce APR	340	400	390
Proactive: Increase Credit Limit	210	140	70

Для визуализации полученных результатов используются метрики, вычисляемые на основе целевых и симулированных фактов. Метрики могут быть предопределенными и пользовательскими (рис. 3). Метрика задается именем метрики (Name), ее типом (Type), который может принимать значения «текущий объем» (Volume) или «кумулятивный объем» (Cumulative), а также формулой для расчета метрики (Expression).

The screenshot displays a software interface for configuring metrics. At the top, there is a table with three columns: 'Name', 'Type', and 'Expression'. Below the table are 'Add ...' and 'Remove' buttons. The 'Metric1' row is highlighted in blue. Below the table is a 'Metric details' section with the following fields:

- Name*: Metric1
- Description: (empty text area)
- Expression*: 2*{Fullfillments Processed}
- Type: Volume (dropdown menu)
- Target comparison: Normal (selected radio button), Reversed (unselected radio button)
- Enabled:

Рис. 3. Конфигурация метрик

В данной модели предопределены следующие метрики:

- общий объем фактов (Volume);
- количество принятых предложений банка клиентами (Number Accepted);
- количество обработанных ответов от клиентов (Responses processed);
- доход (Revenue);
- общее количество выполняемых предложений (Number Fulfilled);
- количество обработанных выполняемых предложений (Fullfillments Processed).

3. Реализация системы сегментирования клиентов

Согласно стандарту IEEE 1472000 «Порядок описания архитектурных решений программных систем, рекомендуемый IEEE» [7] под архитектурой понимается фундаментальное устройство системы, воплощенное в ее компонентах и связях между ними, а также среды и руководящих принципов их дизайна и эволюции.

Отсюда следует, что архитектура программы представляет собой структуру, которая включает элементы программного обеспечения, внешне видимые свойства этих элементов и отношения между ними.

Для работы рассматриваемой модели будем применять вычислительную сеть (грид). На каждом узле этой сети используется база данных Berkeley DB для хранения и обработки исторических и симулированных фактов. Для ускорения выполнения запросов результатов работы с вычислительных узлов применяется кеширование на управляющем узле. Система поддерживает работу с несколькими клиентами одновременно, а также построение нескольких визуализаций результатов как для одной, так и для множества метрик. Более того, пользователи системы могут сохранять выбранные ими представления данных для дальнейшего анализа.

В рассматриваемом случае архитектура системы состоит из следующих основных компонентов (рис. 4):

- вычислительного сервера (App Server);
- управляющего грид-узла (Master Grid Node);
- вычислительных грид-узлов (Grid Nodes);
- хранилища данных (Data store);
- пользовательской системы (User).

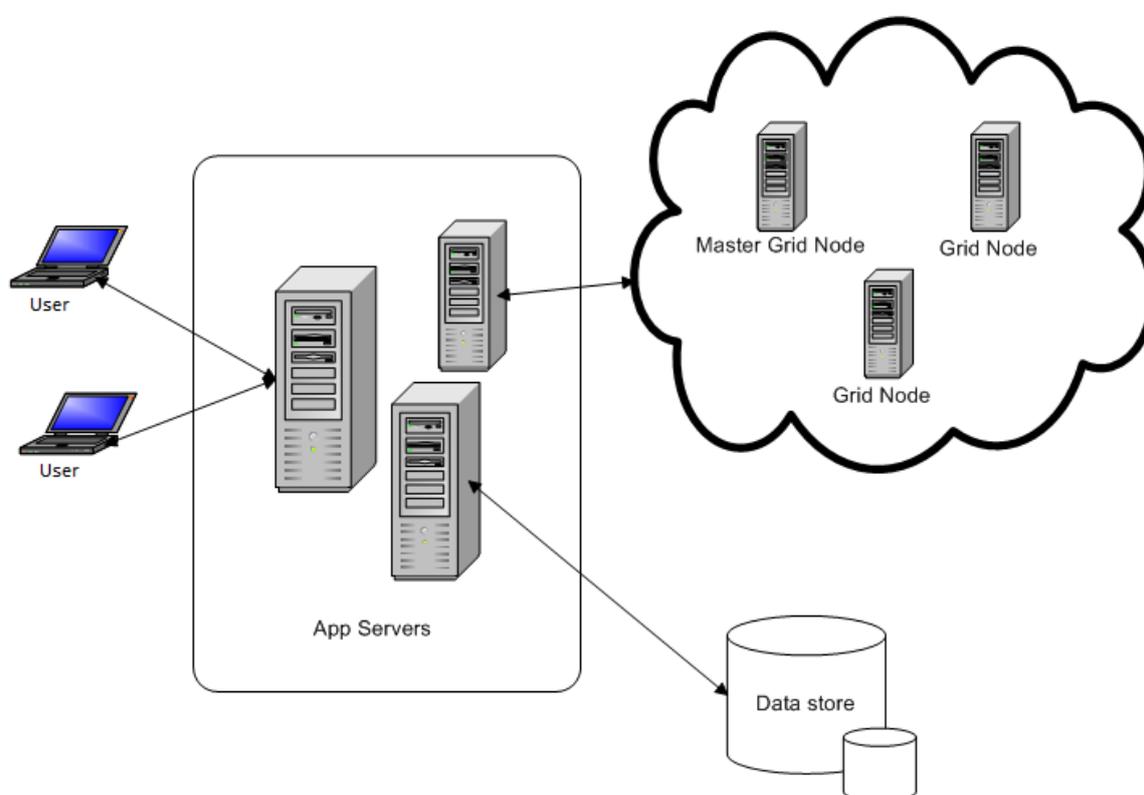


Рис. 4. Архитектура системы сегментирования клиентов

Вычислительный сервер выполняет задачу взаимодействия пользователей и вычислительной сети, а также кеширования данных.

Хранилище данных содержит исторические данные и предоставляет к ним доступ.

Управляющий узел инициирует процесс симуляции, распределяет исторические данные по вычислительным узлам, собирает промежуточные результаты, а также статистику выполнения симуляции с грид-узлов.

На грид-узлах непосредственно выполняется процесс. Пользователь системы является приложением, обеспечивающим возможность работы с параметрами модели, запуска и контроля выполнения симуляции, а также визуализации ее результатов.

Рассмотрим этапы поведения системы в процессе реализации модели (рис. 5). До запуска процесса управляющий узел распространяет на вычислительные узлы документ описания модели на языке XML. Узлы выстраиваются по топологии «кольцо», при этом происходит присвоение каждому узлу уникального ключа, который в процессе симуляции будет использован для идентификации данных, получаемых из вычислительных узлов.



Рис. 5. Процесс реализации модели

Весь процесс симуляции условно можно разбить на четыре этапа. На первом этапе происходит распределение фрагментов данных, участвующих в симуляции между вычислительными узлами. На втором этапе происходит запуск процесса моделирования поведения для получения результатов без изменения исходных параметров. На третьем этапе запускается поиск решений для определения значений фактов с измененными контрольными параметрами системы. Все найденные варианты хранятся в локальной базе вычислительного узла. На последнем этапе управляющий узел запрашивает у каждого узла агрегированные результаты, рассчитываемые по метрикам, которые задаются пользователем модели.

Данные, хранимые на вычислительном узле, описываются двумя структурами – DecisionData и DataFact. DecisionData содержит в себе конкретные значения входных параметров модели, а также возможные значения ее контрольных параметров. Одна запись DecisionData уникальна во всей системе и идентифицируется соответствующим образом. DataFact содержит набор из симулированного и целевого фактов.

В связи с тем что такие наборы данных не зависят друг от друга, их можно обрабатывать параллельно. Благодаря этому достигается атомарность шагов симуляции, в силу чего данный процесс попадает в класс хорошо распараллеливаемых задач.

В основе процесса симуляции лежат генетические алгоритмы. Кроме того, количество целевых и симулируемых фактов прямо пропорционально количеству исторически накопленных фактов и контрольным параметрам системы. В связи с этим для ускорения процесса моделирования целесообразно применить вычислительную сеть, описанную ранее. Атомарность шагов симуляции позволяет получать результаты уже в процессе симуляции, что, в свою очередь, дает возможность пользователю модели анализировать результаты, не дожидаясь окончания процесса. Более того, атомарность операций позволяет вносить изменения в наборы контрольных параметров, не прерывая симуляцию. Новые факты в таком случае будут учтены только на шагах, еще не обработанных к моменту их наступления.

Заключение

Разработанная и предложенная система моделирования поведения клиентов реализована на языке Java 6 применением библиотек параллельного вычисления GridGain. Управляющий сервер располагается на наборе J2EE контейнеров, для хранения информации поддерживается ряд корпоративных баз данных.

На основании сегментации клиентской базы система позволяет определить наиболее прибыльный сегмент клиентов не только в краткосрочной, но и в долгосрочной перспективе.

В отличие от существующих аналогов, таких как решения компаний Oracle и PegaSystems, разработанная система позволяет не только прогнозировать текущее поведение клиентов, но и симулировать поведение с измененными контрольными параметрами. Новизна работы заключается в реализации данной системы с применением алгоритмов симуляции, а также вычислительной сети.

В общем случае решения, следующие из результатов работы предложенной модели, полезны как для оперативного прогнозирования, так и для принятия обоснованных решений в организациях, не преследующих коммерческих целей, в торговых или банковских системах [8].

Список литературы

1. GridGain 3.0: High Performance Cloud Computing [Electronic resource]. – Mode of access : http://www.gridgain.com/media/gridgain_white_paper.pdf. – Date of access : 7.06.2012.
2. Brandic, I. High-level Composition of QoS-aware Grid Workflows: An Approach that Considers Location Affinity / I. Brandic, S. Pllana, S. Benkner [Electronic resource]. – Mode of access : http://www.isi.edu/works06/brandic_works06.pdf. – Date of access : 7.06.2012.
3. Customer Centricity in Insurance Sales, Service and Saving e Customer [Electronic resource]. – Mode of access : <http://www.insurancetech.com/whitepaper/Customer-Insight-Business-Intelligence/Analytics/customer-centricity-in-insurance-sales-service-an-wp1330111991>. – Date of access : 7.06.2012.
4. Peppers, D. Return on Customer: Creating Maximum Value From Your Scarcest Resource / D. Peppers, M. Rogers. – USA : Doubleday, division of random House Inc., 2005. – 304 p.
5. Bass, L. Software Architecture in Practice / L. Bass, P. Clements, R. Kazman. – Second Edition. – USA : Addison Wesley, 2003. – 560 p.
6. Smith S. Managing the Customer Experience: Turning customers into advocates / S. Smith, J. Wheeler. – USA : Financial Times Press, 2002. – 272 p.
7. IEEE Computer Society, IEEE Recommended Practice for Architectural Description of Software-Intensive Systems: IEEE Std 1472000, 2000.
8. Mattison R. Data Warehousing and Data Mining for Telecommunications / R. Mattison. – Norwood, MA : Artech computer Scion Library, 1997. – 237 p.

Поступила 07.06.2012

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

М.К. Bouza, V.S. Gorshunov

MODELING ORGANIZATION CLIENTS BEHAVIOR

A system of modeling organization clients behavior is proposed and studied, as well as its functional and architectural solutions. A client identification mechanism based on control and dynamic parameters of a given objective function is proposed.

УДК 577.29, 51-76

В.В. Баркалин¹, Я.В. Долгая¹, О.А. Козлова²

МОДЕЛИРОВАНИЕ ПРЕДПОЛИМЕРИЗАЦИОННОГО КОМПЛЕКСА ПОЛИМЕРОВ С МОЛЕКУЛЯРНЫМИ ОТПЕЧАТКАМИ ТРИ-О-АЦЕТИЛАДЕНОЗИНА

Рассматривается применение метода компьютерного моделирования QM/MM, совмещающего квантово-механический и молекулярно-механический подходы для моделирования предполимеризационных комплексов молекулы-темплата и мономеров и дальнейшего in-silico исследования биомолекулярных свойств полимеров с молекулярными отпечатками, что позволяет оценить их эффективность и оптимизировать процесс получения и использования при различных условиях. Проводятся моделирование изолированных предполимеризационных комплексов темплата три-о-ацетиладенозина с различным числом мономеров акриламида, метакриловой и мета-винил-бензойной кислот, а также сравнение результатов расчетов методом QM/MM и методом самосогласованного поля с точки зрения компьютерных затрат и параллелизации вычислений.

Введение

Технология молекулярного импринтинга основана на полимеризации функциональных мономеров в присутствии молекул вещества-шаблона (темплатов), которые затем вымываются из полимера с целью получения полостей-отпечатков, соответствующих темплату по форме и наличию функциональных групп. Полученные полимеры используются для выборочного распознавания и извлечения из растворов различных веществ молекул, соответствующих отпечаткам.

Полимеры с молекулярными отпечатками молекулы три-о-ацетиладенозина (ТАА) могут использоваться для получения АТФ-конкурентных ингибиторов, являющихся основным компонентом разрабатываемых лекарственных препаратов, снижающих активность протеинкиназы для лечения воспалений, различных видов рака и диабета. Моделирование изолированного предполимеризационного комплекса показало, что QM/MM-подход позволяет дать сравнительную оценку таким параметрам, как вид функционального мономера и мольное соотношение мономеров к темплатам при относительно небольших вычислительных затратах, что делает возможным применение предложенной методики для исследования влияния растворителей и сшивателей на процесс импринтинга и характеристики полученных полимеров.

Разработка способа производства полимеров с заданными молекулярными отпечатками включает выбор следующих параметров: вида функционального мономера, молекулы которого связываются с темплатом, образуя предполимеризационный комплекс (ППК), и затем полимеризуются; мольного соотношения функционального мономера и темплата, в котором они добавляются в раствор для образования ППК; вида растворителя, в котором происходит образование ППК; мольного отношения сшивающего агента к мономеру (сшивающего коэффициента).

Использование компьютерных технологий и существующих методов моделирования позволит существенно снизить затраты на исследования, проводимые для выбора и оценки указанных параметров, по таким критериям, как энергия связи, пространственная структура ППК и конечного полимера.

Для исследования биомолекулярных систем в настоящее время применяются: методы молекулярной механики, когда все взаимодействия между атомами описываются посредством классических силовых полей; методы, базирующиеся на квантовой механике, а также комбинированный подход квантовой механики и молекулярной (QM/MM). В [1] был предложен способ расчета мольного соотношения темплатов и мономеров и сравнения эффективности мономеров для импринтинга на основе моделирования ППК методом молекулярной механики, который затем уточнялся квантово-механическим полуэмпирическим расчетом с гамильтонианом PM3. При этом основным критерием оценки считалась величина энергии связи, которая рассчитывалась как разница между полной энергией ППК и суммой энергий изолированных

молекул. Следует отметить, что связывание темплата и мономера происходит за счет образования межмолекулярных водородных связей, которые не являются типичными химическими связями, но и не сводятся к простому электростатическому притяжению и, следовательно, не могут быть описаны с помощью классических силовых полей, на которых базируется метод молекулярной механики. Полуэмпирические методы при расчете водородных связей дают заниженную величину для энергии связи и завышенную для длины [2].

Для решения поставленных задач использовался метод QM/MM, в котором вся система разделяется на два региона: квантовый и классический, а энергия каждого региона вычисляется соответственно квантово-механическим и молекулярно-механическим способами. Области с водородными связями включаются в квантовый регион и моделируются методом самосогласованного поля (ССП), а остальная часть системы, включая растворитель, моделируется методом молекулярной механики. Такой подход позволяет с достаточной точностью определить параметры водородных связей и в то же время избежать квантово-механического расчета всей системы. Настоящая работа посвящена исследованию применения QM/MM-подхода для выбора и оценки параметров импринтинга. Для анализа применимости метода проведено моделирование ППК с различными функциональными мономерами (метакриловой кислотой, мета-винил-бензойной кислотой, акриламидом) без учета растворителя. Проведено сравнение результатов с результатами, полученными при моделировании методом СПП. Расчеты произведены с использованием пакета квантово-механических и молекулярно-динамических вычислений NWChem-6.0 на вычислительном грид-сегменте ВУ-ВНТУ Белорусского национального технического университета.

1. Разработка полимеров с молекулярными отпечатками три-о-ацетиладенозина

Технология молекулярного импринтинга [3] включает три этапа. На первом этапе путем смешения в определенной пропорции темплата и мономера в растворителе формируют устойчивые предполимеризационные комплексы из мономеров и темплатов. На втором этапе комплексы полимеризуют в присутствии сшивающего агента. На третьем этапе темплат вымывают из полимера, в результате чего в нем образуются полости-отпечатки для захвата молекул ингибитора.

Исследуется технология производства и эффективность применения полимеров с использованием в качестве темплата аналога АТФ – три-о-ацетиладенозина. В качестве возможных функциональных мономеров рассматриваются молекулы метакриловой кислоты (МАК), мета-винил-бензойной кислоты (МВБ) и акриламида (АКР), из которых необходимо выбрать наиболее подходящую. На рис. 1 схематично изображен процесс импринтинга ТАА и последующего применения полученного полимера для селективного извлечения ингибиторов.

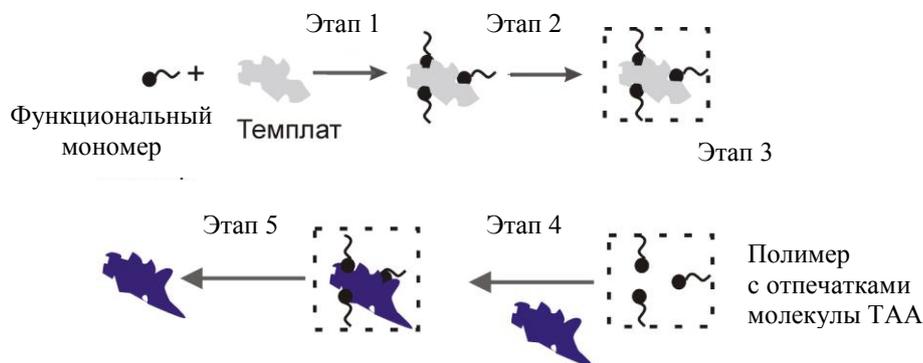


Рис. 1. Этапы процесса получения и применения полимера с отпечатками молекулы ТАА: 1 – получение ППК; 2 – полимеризация; 3 – вымывание темплата из полимера; 4 – селективное извлечение молекул, которые соответствуют сформированным отпечаткам по форме, размеру и наличию функциональных групп, способных взаимодействовать с функциональными группами мономеров; 5 – вымывание выбранных молекул

2. Методика применения компьютерного моделирования для решения задачи оценки условий импринтинга

2.1. Критерии оценки параметров

Предлагаемая методика включает моделирование каждого из этапов импринтинга и применения полимеров (см. рис. 1).

При моделировании первого этапа (процесса получения ППК) производится оценка условий импринтинга по следующим критериям: пространственная структура ППК, энергия связи ППК.

Так как пространственная структура ППК определяет вид полости (отпечатка), которая должна имитировать участок связывания АТФ протеинкиназы, мономеры в ППК должны образовывать связи только с аденином АТФ. Таким образом, структуры характеризуются наличием водородных связей «атом 1 – атом 2», где атом 1 – атом темплата, атом 2 – атом мономера, между которыми образуется водородная связь за счет атома водорода, связанного с одним из атомов ковалентной связью. Для определения наиболее вероятной пространственной структуры ППК при заданных параметрах необходимо провести расчет энергий всех возможных структур и выбрать структуру, соответствующую наименьшей энергии. Для этого необходимо задавать приближенные начальные координаты атомов ППК, располагая молекулы мономеров вокруг темплата так, чтобы между ними могли образовываться водородные связи. Мольное соотношение определяет количество функциональных мономеров, присоединенных к темплату. Молекулы растворителя равномерно распределяются в ячейке моделирования с учетом плотности раствора. Выполняя оптимизацию геометрии, можно найти локальный минимум энергии, который будет соответствовать энергии заданной структуры.

Энергия связи ППК вычисляется как разность суммы энергий, составляющих ППК молекул, и полной энергии ППК и характеризует устойчивость ППК:

$$E_{св} = E_{ППК} - (E_{ТАА} + \sum E_{мон}). \quad (1)$$

При моделировании второго этапа (процесса полимеризации) определяется сшивающий коэффициент. В систему добавляются дополнительные молекулы мономеров и сшивателей, количественное соотношение которых определяется сшивающим коэффициентом, для которого производится оценка. Затем проводится оптимизация геометрии. Критериями оценки являются пространственная структура полимера, которая включает форму, пористость полимерной матрицы и доступность связывающих полостей, а также энергия связи образованного полимера, определяемая как разница между энергией полученного полимера и энергией молекул, его составляющих.

На третьем этапе полученная полимерная матрица, содержащая темплаты, помещается в растворитель для отмывки. Моделирование этого этапа проводится методом молекулярной динамики, в результате которого определяется, сколько темплатов вымыто из полостей-отпечатков. Таким образом, на втором и третьем этапах определяется селективная поверхность полимера.

На последнем этапе проводится анализ способности полимера к распознаванию молекул-ингибиторов. Для этого осуществляется оптимизация геометрии полимера с отпечатками, в котором на место молекул-шаблонов в полость помещены молекулы-ингибиторы. Затем определяется энергия связи полимера с молекулами-ингибиторами, величина которой характеризует аффинность полостей-отпечатков.

Комплексная оценка результатов, полученных на всех этапах моделирования, позволяет выбрать наиболее подходящие для импринтинга функциональный мономер, растворитель, мольные соотношения мономеров к темплату и сшивателю.

2.2. Применение метода QM/MM для оптимизации геометрии и вычисления энергии связи

При моделировании всех этапов импринтинга используется метод QM/MM. В квантовый регион включаются все атомы темплата (рис. 2), а также атомы мономеров, участвующие в

образовании водородных связей и образующие с ними связь (рис. 3), в классический – оставшиеся атомы мономеров, сшивателей и молекулы растворителя.

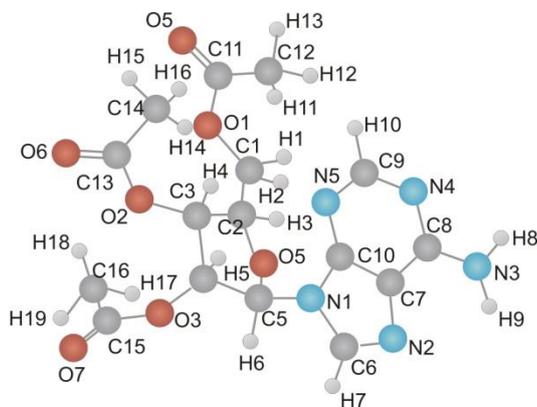


Рис. 2. Молекула ТАА

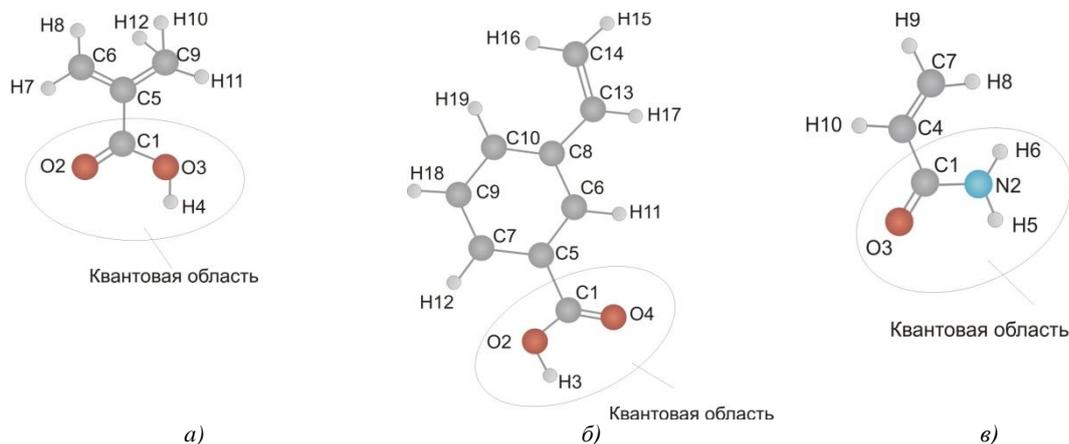


Рис. 3. Функциональные мономеры: а) метакриловая кислота; б) мета-винил-бензойная кислота; в) акриламид

Полная энергия системы рассчитывается как сумма энергий классической и квантовой областей, а также энергии взаимодействия между ними.

Алгоритм оптимизации геометрии в пакете NWChem включает поочередную минимизацию энергий классического и квантового регионов методом квази-Ньютона с ограниченной памятью до тех пор, пока не будет достигнут критерий сходимости. Молекулярные связи, пересекающие регионы, заменяются связями с атомом водорода (атомом связи), место положения которого вычисляется следующим образом:

$$R_{св} = (1 - g) \cdot R_{квант} + g \cdot R_{класс}, \quad (2)$$

где g – масштабирующий множитель, равный 0,709; $R_{квант}$, $R_{класс}$ – радиус-векторы атомов, участвующих в этой связи, с квантовой и классической сторон соответственно [4].

Сначала оптимизируется квантовый регион вместе с атомами связи и вычисляется его электростатическое поле. При оптимизации классического региона атомы, включенные в квантовый регион, и атомы связи считаются неподвижными и представляются в виде электростатического потенциала. Цикл повторяется до достижения критерия сходимости.

Выбор базиса для расчета энергии квантового региона определяется наличием в системе водородных связей и затратами процессорного времени. Неэмпирические расчеты для систем с водородными связями в настоящее время чаще всего проводят с использованием валентно-расщепленных базисов (обычно 3-21G или 4-31G). Их выбор связан с необходимостью идти на компромисс между точностью расчета и затратами машинного времени. Расчеты в этих базисах намного завышают энергию образования водородных связей, занижают их длину (расстояние между гетероатомами), а также существенно переоценивают стабильность циклических структур с водородными связями. Включение в базис поляризационных орбиталей устраняет эти недостатки. Таким образом, при использовании неэмпирических методов расчеты для систем с водородными связями желательно проводить в базисах с поляризационными орбиталями, т. е. использовать базис 6-31G* или 6-31G** [5], что практически трудно осуществить для моделирования систем органической химии и биологии, имеющих обычно большие размеры, квантово-механическими методами. Однако использование данных базисов возможно при применении QM/MM-подхода.

Авторами проведено исследование применимости неполяризованного и поляризованного базисов в моделировании водородных связей на примере расчета энергий связи между двумя молекулами воды. Рассчитанная энергия водородной связи с использованием базиса 6-31G равна 32,81 кДж/моль, с использованием базиса 6-31G** – 23,05 кДж/моль. Экспериментальное значение энергии связи – 21 кДж/моль. В этой связи для достижения большей точности расчетов был выбран поляризованный базис 6-31G** для квантового региона. Для молекулярно-механической части использовалось силовое поле *amber*, которое успешно применяется для моделирования биомолекулярных систем [6].

3. QM/MM-моделирование предполимеризационного комплекса

На рис. 4 показаны места присоединения мономеров при моделировании и обозначены водородные связи.

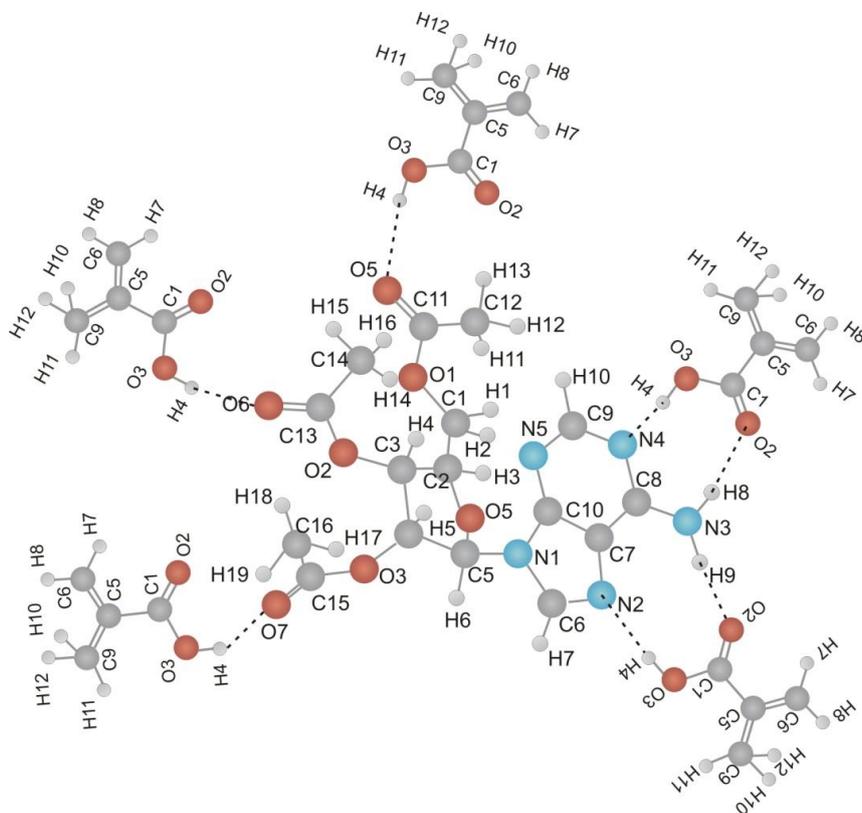


Рис. 4. Молекула ТАА с пятью присоединенными мономерами метакриловой кислоты

Моделирование проводилось для структур с одним – пятью мономерами, присоединенными к темплату.

В табл. 1 представлены результаты вычислений энергий связи методом QM/MM для комплексов с представленными мономерами, присоединенными к различным атомам темплата (местам присоединения).

Таблица 1

Энергии связи (*E_{св}*) ППК, кДж/моль

Кол-во мономеров	Места присоединения	<i>E_{св}</i> МАК	<i>E_{св}</i> МВБ	<i>E_{св}</i> АКР
1	N4, N3	-66,69	-58,5	-46,00
1	N2, N3	-64,23	-54,21	-44,44
1	O7	-49,77	-43,58	-33,03
1	O6	-32,45	-41,05	-32,91
1	O5	-47,96	-43,82	-10,75
2	N4, N3, N2, N3	-126,88	-110,39	-80,29
3	N4, N3, N2, N3, O7	-174,88	-149,63	-110,50
3	N4, N3, N2, N3, O6	-172,57	-147,36	-110,39
3	N4, N3, N2, N3, O5	-171,58	-148,87	-111,31
4	N4, N3, N2, N3, O6, O7	-220,73	-187,15	-139,71
4	N4, N3, N2, N3, O5, O6	-224,26	-186,47	-142,09
4	N4, N3, N2, N3, O5, O7	-221,14	-186,59	-140,96
5	N4, N3, N2, N3, O5, O6, O7	-260,55	-224,04	-169,67

4. Сравнение результатов QM/MM- и ССП-подходов молекулярно-механического моделирования

Для проверки результатов методом ССП были вычислены энергии оптимизированных методом QM/MM систем в базисе 6-31 G** и рассчитаны энергии связей (табл. 2).

Таблица 2

Энергии связи (*E_{св}*) ППК, оптимизированных в QM/MM, рассчитанные методом ССП, кДж/моль

Кол-во мономеров	Места присоединения	<i>E_{св}</i> МАК	<i>E_{св}</i> МВБ	<i>E_{св}</i> АКР
1	N4, N3	-64,02	-57,13	-44,75
2	N4, N3, N2, N3	-121,45	-107,47	-78,07
3	N4, N3, N2, N3, O7	-166,11	-144,93	-107,01
4	N4, N3, N2, N3, O6, O7	-210,28	-180,58	-136,08
5	N4, N3, N2, N3, O5, O6, O7	-246,80	-216,07	-162,95

Также была проведена оптимизация комплексов, мономеров и темплата методом ССП в базисе 6-31 G**. Результаты вычисленных по формуле (1) энергий связей приведены в табл. 3.

Таблица 3

Энергии связи ППК, рассчитанные методом ССП, кДж/моль

Кол-во мономеров	Места присоединения	<i>E_{св}</i> МАА	<i>E_{св}</i> МВБ	<i>E_{св}</i> АКР
1	N4, N3	-64,61	-59,68	-50,75
2	N4, N3, N2, N3	-123,84	-108,78	-85,71
3	N4, N3, N2, N3, O7	-171,89	-147,34	-116,76
4	N4, N3, N2, N3, O6, O7	-223,93	-185,5	-148,16
5	N4, N3, N2, N3, O5, O6, O7	-269,99	-197,61	-

Методом молекулярной механики для ППК с метакриловой кислотой были вычислены энергии связей (табл. 4). Дальнейшие исследования не проводились, так как из полученных результатов следует невозможность образования ППК с двумя мономерами, что противоречит результатам ССП-моделирования, которое является более точным, и экспериментальным данным [7].

Таблица 4

Энергии связи ППК, рассчитанные методом молекулярной механики, кДж/моль

Кол-во мономеров	Места присоединения	Есв МАА	Есв МВБ	Есв АКР
1	N4, N3	-64,61	-59,68	-50,75
2	N4, N3, N2, N3	-123,84	-108,78	-85,71
3	N4, N3, N2, N3, O7	-171,89	-147,34	-116,76
4	N4, N3, N2, N3, O6, O7	-223,93	-185,5	-148,16
5	N4, N3, N2, N3, O5, O6, O7	-269,99	-197,61	-

На рис. 5 изображены графики зависимости энергий связи от числа функциональных мономеров в ППК, рассчитанных методами ССП и QM/MM.

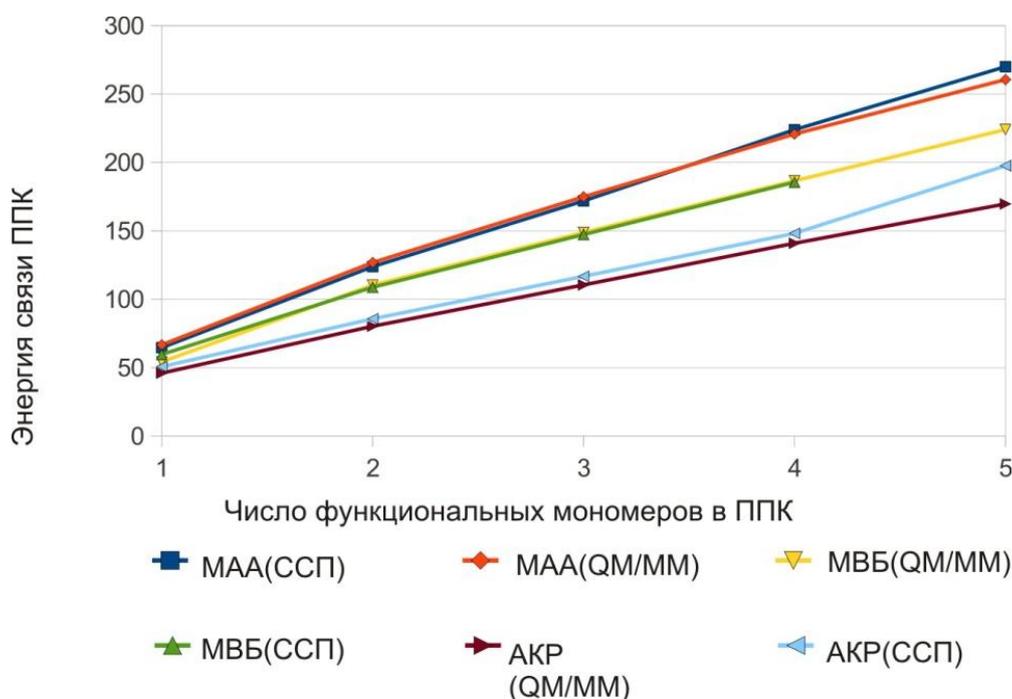


Рис. 5. Графики зависимости энергии связи ППК от числа мономеров

5. Сравнение вычислительной эффективности применения QM/MM-алгоритма, реализованного в NWChem для молекулярного импринтинга

Рассмотрим графики на рис. 6 и 7. Видно, что время вычисления энергии методом QM/MM практически не меняется с увеличением ППК, а время вычисления методом ССП возрастает с разной скоростью. Время оптимизации геометрии не всегда возрастает, так как количество шагов оптимизации зависит от близости исходной системы к конечной, однако при количестве атомов, большем 95, заметным становится преимущество метода QM/MM. Это связано с тем, что время расчета энергии системы заметно возрастает для ССП-метода и оказывается решающим фактором даже при меньшем числе шагов ССП-метода по сравнению с QM/MM.

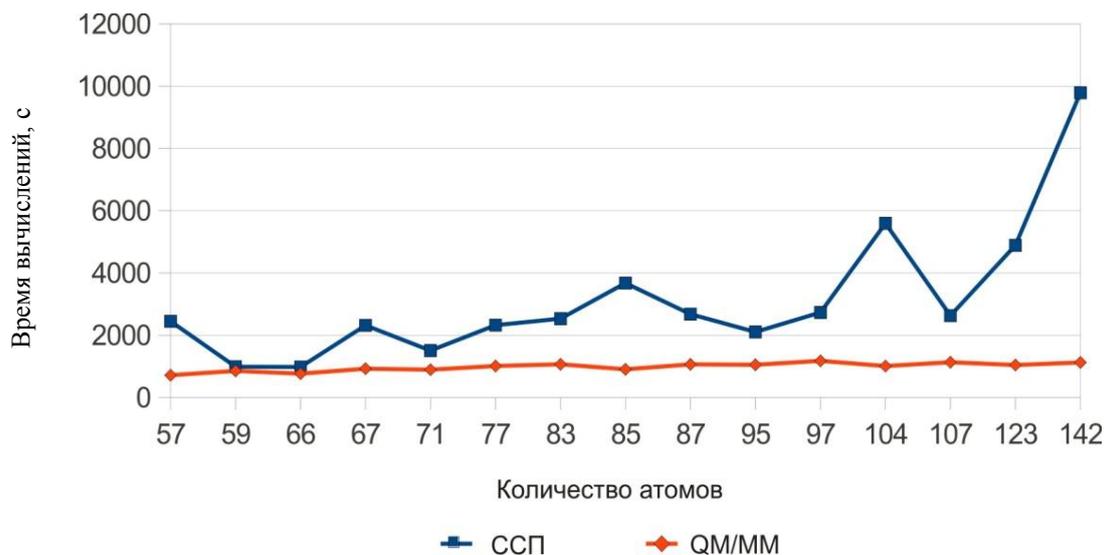


Рис. 6. График времени расчета оптимизации геометрии некоторых ППК методами ССП и QM/MM на восьми ЦПУ (Intel Xeon E5520, 2.27GHz)

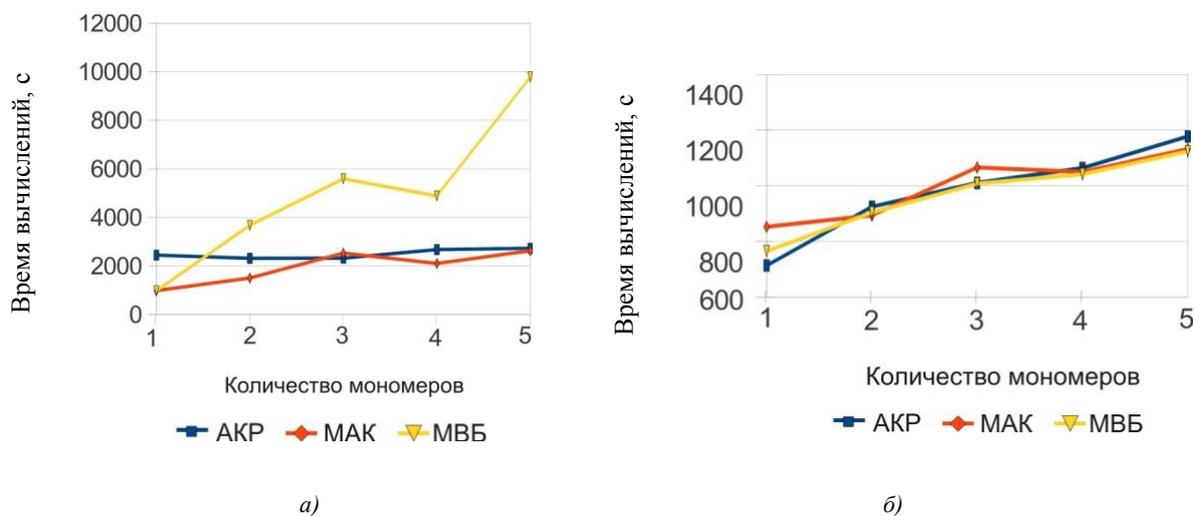


Рис. 7. График времени расчета энергии ППК в зависимости от числа мономеров: а) для метода ССП; б) для метода QM/MM

Для оценки параллелизуемости метода QM/MM были проведены расчеты энергии ППК с пятью функциональными мономерами мета-винил-бензойной кислоты на количестве вычислительных ядер от 4 до 40. Так как один вычислительный узел содержит восемь ядер, число ЦПУ более восьми достигается увеличением количества вычислительных узлов. Из графика на рис. 8 видно, что увеличение числа вычислительных узлов не ведет к ускорению вычислений в отличие от увеличения числа ядер в пределах одного вычислительного узла.

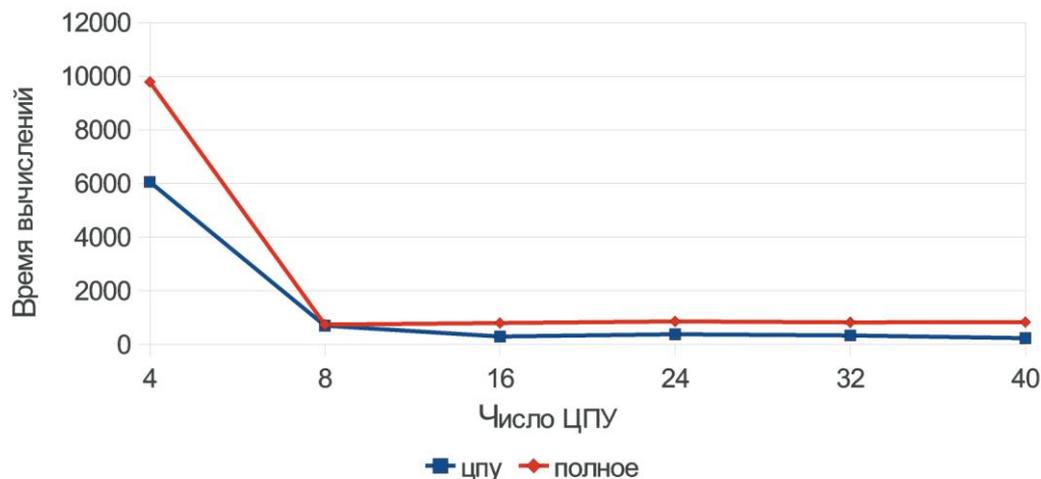


Рис. 8. Зависимость времени расчета энергии ППК (ТАА и 5 ТВБ) от числа ЦПУ

Заключение

На основании вычисленных методом QM/MM значений энергий образования ППК можно сделать вывод, что наилучшим образом связываются с ТАА молекулы метакриловой кислоты, затем мета-винил-бензойной, а наиболее слабые связи образуются с акриламидом. Энергия образования ППК возрастает с увеличением количества присоединенных к темплату мономеров. Вычисления методом ССП приводят к таким же результатам. При расчете небольших систем метод QM/MM не дает значительного выигрыша по сравнению с методом ССП. Это обусловлено тем, что в квантовую область включено больше атомов, чем в классическую. Кроме того, значительное влияние на скорость сходимости решения также оказывают начальные условия, поэтому отсутствует прямая зависимость между количеством атомов в системе и временем оптимизации. Однако уже для системы с пятью молекулами мета-винил-бензойной кислоты, состоящей из 142 атомов, время расчета методом QM/MM значительно сокращается по сравнению с методом ССП. При дальнейшем увеличении системы (добавлении растворителя и полимера) становится затруднительно использовать метод ССП из-за слишком большого времени расчета, а при использовании метода QM/MM молекулы будут добавлены только в классический регион. Это не окажет существенного влияния на время моделирования, так как взаимодействие между молекулами полимера и растворителя будет происходить за счет ковалентных связей и электростатических сил. Предполагается, что применение QM/MM-моделирования для оценки параметров импринтинга ТАА повысит эффективность и качество процесса разработки технологии получения полимеров с отпечатками ТАА.

Работа выполнена в рамках задания 6.06 ГПНИ «Информатика и космос» 2011–2015 гг. Авторы выражают благодарность соисполнителям задания В.В. Нелаеву (БГУИР), С.В. Медведеву (ОИПИ НАН Беларуси), А.В. Никинину (ГрГУ) за плодотворные дискуссии и доступ к компьютерным ресурсам.

Список литературы

1. Farrington, K. Predicting the performance of molecularly imprinted polymers: Selective extraction of caffeine by molecularly imprinted solid phase extraction / K. Farrington, E. Magner, F. Regan // *Analytica Chimica Acta*. – 2006. – Vol. 566. – P. 60–68.
2. Jurema, M.W. Ability of the PM3 Quantum-Mechanical Method to Model Intermolecular Hydrogen Bonding between Neutral Molecules / M.W. Jurema, G.C. Shields // *Journal of Computational Chemistry*. – 1993. – Vol. 14, № 1. – P. 89–104.

3. Lakka, A. Fabrication and characterization of tri-oacetiladenosine imprinted polymers for screening and isolation of protein kinase inhibitors / A. Lakka // Biomaterials and bionanomaterial, may 2-9 2010. – Iraklion, Crete: Greece Book of Abstract, 2010. – P. 28.
4. NWChem user support [Electronic resource]. – Mode of access : http://www.nwchem-sw.org/index.php/QMMM_Parameters. – Date of access: 01.06.2012.
5. Бурштейн, К.Я. Квантово-механические расчеты в органической химии и молекулярной спектроскопии / К.Я. Бурштейн, П.П. Шорыгин. – М. : Наука, 1989. – 98 с.
6. Senn, H.M. QM/MM Methods for Biomolecular Systems / H.M. Senn, W. Thiel // Angew. Chem. Int. Ed. – 2009. – Vol. 48. – P. 1198–1229.
7. Lakka, A. Application of molecularly imprinted polymers (MIP) in drug discovery inhibitors / A. Lakka, A. Tsakalof // Biomaterials and bionanomaterial, may 2-9 2010. – Iraklion, Crete, Greece Book of Abstract., 2010. – P. 28.

Поступила 17.01.2013

¹Белорусский национальный
технический университет,
Минск, Независимости, 65
e-mail: douhaja_grid@bntu.by

²Белорусский государственный
университет информатики
и радиоэлектроники,
Минск, ул. П. Бровки, 6
e-mail: olga.bsuir@gmail.com

V.V. Barkaline, Y.V. Douhaya, O.A. Kozlova

QM/MM MODELING OF TAA IMPRINTED POLYMERS PRE-POLYMERIZATION COMPLEX

A QM/MM method for pre-polymerization complex computer modeling followed by an in-silica research of bimolecular properties of imprinted polymers can be applied to optimize and estimate the efficiency of imprinted polymers fabrication and usage under various conditions. Isolated pre-polymerization complex modeling of 3-o-acetiladenosine with various number of monomers of acrilamid, metacrilic acid and 3-vinyl benzol acid and with different molar ratio is carried out. The results of QM/MM modeling are compared with the SCF method results in terms of computer resource consumption and parallelization efficiency.

УДК 004.7:004.93:004.942

В.Г. Родченко, Е.В. Олизарович, А.И. Жукевич

ПРИМЕНЕНИЕ МЕТОДОВ РАСПОЗНАВАНИЯ ДЛЯ ПОСТРОЕНИЯ СИСТЕМ ДИАГНОСТИКИ КОМПЬЮТЕРНЫХ СЕТЕЙ

Предлагается универсальный метод диагностики компьютерных сетей, базирующийся на использовании математического аппарата теории распознавания образов. Представляется способ формализации состояния сети на основе наблюдаемых характеристик трафика. Рассматриваются механизмы реализации всех основных этапов диагностики, связанных с формированием словарей признаков, выполнением процедур обучения и классификации.

Введение

Компьютерные сети (КС) как объекты исследования имеют выраженные свойства сложных систем, в которых каждый компонент оказывает влияние на функционирование других. Современное состояние и тенденции развития телекоммуникационных сетей характеризуются распространением технологий мультисервисного доступа, беспроводных коммуникаций, виртуальных сетей, внедрением новых протоколов передачи данных и технологий распределенной обработки информации, повышением уровня компетентности пользователей. Эти процессы вызывают постоянное обновление сетевых устройств, программного обеспечения и характеристик сетевого трафика, что обуславливает появление качественно новых диагностических задач и приводит к необходимости постоянной модернизации методов и средств диагностики.

Понятия технической диагностики КС, определенные в нормативных документах, а также существующие технические средства и методы диагностики ориентированы на измерение и оценку значений среднестатистических показателей или точных значений (сигнатур), номенклатура которых постоянно возрастает и становится избыточной для конкретных задач управления. При этом анализ показателей возлагается на персонал, и результат существенно зависит от опыта и квалификации имеющихся специалистов. В целом средства диагностики КС ориентированы на определенный тип инцидентов, оборудования и программного обеспечения, что недостаточно в современных условиях.

Методы моделирования телетрафика, используемые при проектировании и теоретическом исследовании КС, также требуют дополнения, поскольку они, как правило, не могут непосредственно применяться для практической диагностики ввиду сложности их построения и невозможности учета всех необходимых данных. В связи с этим перспективным направлением является изучение динамически изменяющихся закономерностей функционирования телекоммуникационных сетей на основе апостериорного анализа свойств трафика гетерогенных КС с целью поиска новых методов моделирования и средств управления.

Статья посвящена описанию универсального метода диагностики КС, который основан на использовании математического аппарата распознавания образов. Реализованные механизмы диагностики инвариантны относительно технической структуры сети и ориентированы на автоматизацию операций мониторинга параметров информационных потоков и классификации состояния.

1. Постановка задачи

Комплексное решение задачи диагностики предполагает выполнение трех основных этапов, связанных, во-первых, с формализацией диагностируемых состояний, во-вторых, с реализацией процедуры обучения при построении образов эталонов диагностируемых состояний и, в-третьих, с постановкой заключительного диагноза, т. е. с классификацией состояния.

На первом этапе должны быть определены множество диагностируемых состояний (алфавит классов), а также набор наблюдаемых характеристик трафика и способы их формального

представления в виде априорного словаря признаков. В результате выполнения этого этапа должна быть сформирована классифицированная обучающая выборка (КОВ).

Второй этап связан с построением формальных образов эталонов состояний, компактных и разделенных в многомерном признаковом пространстве принятия решений. На основе анализа содержимого обучающей выборки выполняется процедура сепарирования признаков по степени их информативности с точки зрения разделения образов эталонов состояний, формируются уточненный словарь признаков и соответствующее пространство принятия решений, в котором размещаются образы эталонов состояний.

Заключительный этап предполагает выполнение процедуры классификации путем сопоставления образа наблюдаемого состояния КС и сформированного на предыдущем этапе множества образов эталонов состояний.

Распространенные средства диагностики КС обычно ориентированы только на решение третьего или, реже, второго этапа в виде программно-аппаратных комплексов, которые не могут быть модернизированы применительно к условиям конкретной КС. Такое положение позволяет эффективно решать распространенные известные диагностические задачи и исключать ошибки персонала, однако имеет и ряд недостатков:

1. Эталоны состояний, как правило, представляют собой одномерные величины, отражающие состояние одного фактора без анализа его взаимосвязи с другими, что не позволяет в полной мере осуществлять анализ состояния и прогноз развития сложных сетевых процессов.

2. Применение теоретически рассчитанных пороговых значений типовых параметров далеко не всегда позволяет оптимизировать структуру сложных сетей для решения специализированных задач диагностики в условиях конкретной сети.

3. Отсутствует возможность оперативного решения нестандартных эксплуатационных задач, анализа новых угроз, обнаружения аномальных состояний.

Таким образом, сегодня в процессе эксплуатации КС требуется не только оценивать соответствие известных параметров типовым значениям, но и проводить *анализ работы КС с целью обнаружения новых факторов или динамики изменения эталонных значений*. Существующие диагностические средства не предусматривают возможности решения таких задач.

Авторами разработан универсальный метод исследования и диагностики КС на основе анализа наблюдаемых характеристик трафика с использованием аппарата теории распознавания образов. Общая схема решения задачи диагностики состояний компьютерной сети формально может быть представлена в виде последовательности преобразований

$$S \longrightarrow C \longrightarrow A \longrightarrow T \longrightarrow A^* \longrightarrow E \longrightarrow R, \quad (1)$$

где S – множество классов диагностируемых состояний, C – словарь наблюдаемых (измеримых) характеристик сети, A – априорный словарь признаков, T – классифицированная обучающая выборка, A^* – уточненный словарь признаков для построения пространства решений, E – множество эталонов диагностируемых классов состояний, R – множество решений.

Для реализации данной схемы в рамках настоящей работы исследованию подвергалась задача разработки и анализа необходимых для выполнения преобразований (1) алгоритмов:

- формализации состояния сети и построения априорного словаря признаков;
- формирования классифицированной обучающей выборки;
- сепарирования признаков из априорного словаря по степени их информативности для построения пространства решений;
- построения и представления образов эталонов диагностируемых состояний в пространстве решений;
- постановки заключительного диагноза.

Основой построения моделей и методов является применение аппарата теории распознавания образов и интеллектуального анализа данных [1–5].

2. Формализация состояний компьютерной сети

Процессы передачи данных, изначально инициированные прикладными программами, реализуются в виде потока пакетов в среде передачи. В настоящей работе принимается, что трафик КС как совокупность всех передаваемых пакетов является отражением технических и потребительских характеристик изучаемого сегмента сети. При исследовании различных аспектов работы КС под трафиком может пониматься: упорядоченный во времени поток кадров, поступающих на аппаратный порт; поток пакетов, проходящих через логический интерфейс; совокупность сообщений, полученных и отправленных прикладной программой.

В общем случае, поскольку передача данных в КС представляет собой многоуровневый процесс, каждое передаваемое по сети сообщение может быть формально описано в виде совокупности полей заголовков сетевых протоколов. Заголовки представляют собой форматированные области данных, включающие сведения об адресах источника и получателя сообщения, формате данных, длине сообщения, состоянии сеанса и т. д. Состав и структура информации, содержащейся в заголовках каждого уровня, известны и регламентируются спецификациями применяемых протоколов. Каждый передаваемый в КС кадр содержит информацию о всех сеансах, задействованных на разных уровнях в процессе взаимодействия клиента с сервером. Поэтому любой транспортный пакет, переданный через сегмент КС, можно описать конечным множеством стандартизированных параметров $\{x_1^{(1)}, \dots, x_{j(i)}^{(i)}\}$, где $x_{j(i)}^{(i)}$ – значение j -го поля заголовка, предусмотренного на i -м уровне эталонной модели OSI/ISO. В предельном случае параметры $x_{j(i)}^{(i)}$ составляют универсальное множество – генеральную совокупность характеристик пакетов, включающую все возможные поля заголовков. Тогда каждый пакет данных в КС независимо от способа организации измерений может быть описан в виде вектора наблюдаемых характеристик:

$$\mathbf{x} = (x_1, \dots, x_p), \quad (2)$$

где x_i – наблюдаемое (измеренное) значение характеристики $c_i \in C$, $i = \overline{1, p}$; p – количество характеристик кадра, измеримых программно-аппаратными средствами.

Для решения задачи диагностики целесообразно из всех возможных полей заголовка пакета выбрать только информативные, т. е. влияющие на качество классификации. Оценка информативности выполняется в два этапа:

1) *формализация* – эвристический выбор наиболее перспективных наблюдаемых (измеримых) параметров и построение на их основе априорного словаря для формального описания трафика;

2) *анализ* – исследование элементов априорного словаря на основе методов прикладной статистики.

Процессы передачи пакетов в реальной сети имеют, как правило, асинхронный и недетерминированный характер. Время возникновения пакета в канале передачи определяется как законами функционирования прикладных систем, обычно неизвестными наблюдателю, так и работой механизма доступа к среде передачи, результат работы которого в нагруженной сети также не может быть точно рассчитан. Таким образом, каждый наблюдаемый в сети пакет, формализованный в виде вектора \mathbf{x} , содержит информацию только об одном сеансе передачи и недостаточно информативен для диагностики состояния КС в целом.

Для практического применения данных, содержащихся в заголовках пакетов, требуется построить агрегированное описание трафика на основе фильтрации и селективного усиления значений наблюдаемых характеристик пакетов с целью получения формальных признаков состояния A . Для этого предлагается провести предварительную обработку данных о пакетах, зарегистрированных за определенный период измерений:

$$\mathbf{y} = (y_1, \dots, y_n) = f(\mathbf{x}(t), \Delta t_x), \quad (3)$$

где y_j – значение признака состояния сети $a_j \in A, j = \overline{1, n}$; n – количество используемых для диагностики признаков; Δt_x – интервал агрегации; $\mathbf{x}(t)$ – значение вектора наблюдаемых характеристик, $t = [t_0, \Delta t_x)$, t_0 – начало интервала агрегации.

Величина Δt_x зависит от особенностей задачи и должна соответствовать интервалу наблюдения, в течение которого сеть проявляет все значимые для диагностики свойства. Для КС, находящейся в стационарном состоянии, результат диагностики для любого интервала Δt_x должен быть одинаков независимо от начала отсчета.

Длительность интервала Δt_x определяет оперативность выполнения диагностики, поэтому должна быть минимально достаточной для проявления трафиком своих свойств. Определение величины Δt_x может выполняться следующим образом: выбираться из типовых справочников, формироваться на основе требований диагностической задачи, выбираться на основе экспертных оценок, рассчитываться на основе предварительного анализа.

Словарь признаков $A = \{A_1, A_2, \dots, A_n\}$ и соответствующие значения координат вектора \mathbf{y} могут формироваться на основе следующих типовых преобразований [6]:

1. Агрегация и фильтрация:

- суммирование значений координаты x_i векторов \mathbf{x} , измеренных за период Δt_x :

$$y_j(\Delta t_x) = \sum_{q=1}^m x_{iq}, \quad (4)$$

где $j = \overline{1, n}$ – идентификатор признака состояния a_j ; m – количество кадров, измеренных за период Δt_x ; i – идентификатор используемой характеристики кадра c_i ;

- подсчет количества событий вида $x_i = B$, зафиксированных за период Δt_x :

$$y_j(\Delta t_x) = \sum_{q=1}^m r, \text{ где } \begin{cases} r = 1 \text{ при } x_i = B; \\ r = 0 \text{ при } x_i \neq B, \end{cases} \quad (5)$$

где $j = \overline{1, n}$ – идентификатор признака состояния a_j ; m – количество кадров, измеренных за период Δt_x ; i – идентификатор используемой характеристики кадра c_i ; B – заданная сигнатура;

- подсчет количества зафиксированных за период Δt_x случаев одновременного наступления событий $x_i = B_1$ и $x_k = B_2, i \neq k$:

$$y_j(\Delta t_x) = \sum_{q=1}^m r, \text{ где } \begin{cases} r = 1 \text{ при } x_i = B_1 \wedge x_k = B_2; \\ r = 0 \text{ при } x_i \neq B_1 \vee x_k \neq B_2, \end{cases} \quad (6)$$

где $j = \overline{1, n}$ – идентификатор признака состояния a_j ; m – количество кадров, измеренных за период Δt_x ; i, k – идентификаторы используемых характеристик кадра c_i, c_k ; B_1, B_2 – заданные сигнатуры.

2. Расчет статистических характеристик выборки k результатов измерений за период Δt_x :

- выборочное среднее значение:

$$y_j(\Delta t_x) = \bar{x}_i = \frac{1}{k} \sum_{i=1}^k x_i; \quad (7)$$

- выборочная дисперсия:

$$y_j(\Delta t_x) = s^2 = \frac{1}{k} \sum_{i=1}^k x_i^2 - \left(\frac{1}{k} \sum_{i=1}^k x_i \right)^2. \quad (8)$$

Приведение к единому масштабу полученных значений параметров y_j обеспечивается нормировкой по диапазону разброса значений:

$$y_j^* = \frac{y_j - \min_{j=1,m}(y_j)}{\max_{j=1,m}(y_j) - \min_{j=1,m}(y_j)}, \quad (9)$$

где $j = \overline{1, n}$ – идентификатор признака состояния; m – объем выборки результатов измерений.

Вектор значений признаков \mathbf{y} , полученный на основе вектора первичных характеристик \mathbf{x} с использованием преобразований (4) – (9), представляет собой *формализованное описание сетевого трафика*, которое может быть использовано для анализа и диагностики компьютерных сетей. Применение такой формализации соответствует поставленным задачам и позволяет построить универсальный метод диагностики со следующими свойствами:

- независимость от средств измерений и возможность синтеза признаков на основе разнородных наблюдаемых данных;
- возможность получения новых знаний о свойствах КС на основе признаков, которые недоступны для методов прямых измерений;
- возможность предварительного преобразования первичных данных любого типа в порядковые величины, пригодные для сравнения и статистической обработки.

Предложенный в рамках настоящей статьи алгоритм формализации описания сетевого трафика (3) рассматривается как основа для построения *модели состояния сети* с использованием образов эталонных и диагностируемых состояний КС.

3. Построение пространства принятия решений и образов эталонов

Исследование и диагностика КС на основе использования аппарата теории распознавания и представленного выше алгоритма формализации предполагает решение следующих задач:

1. Формирование набора из k диагностируемых состояний и априорного словаря из n признаков состояния КС $A = \{A_1, A_2, \dots, A_n\}$. Первичное формирование априорного словаря A может осуществляться либо на основе типовых библиотек, созданных на основе опыта решения подобных задач, либо на основе предположений экспертов.

2. Анализ признаков из априорного словаря и построение пространства принятия решений A^* для исключения неинформативных и неразделяющих признаков. Оценка информативности признаков априорного словаря выполняется с использованием классифицированной обучающей выборки T .

3. Построение множества многомерных эталонов состояния КС $E = \{E_1, E_2, \dots, E_k\}$, где k – количество диагностируемых состояний.

4. Классификация исследуемого состояния в пространстве решений R .

Результативность предлагаемого метода диагностики КС возрастает с увеличением размерности априорного словаря признаков, поэтому на этапе формирования этого словаря необходимо стремиться найти наибольшее количество характеристик сети, логически связанных с диагностируемыми состояниями. Однако для повышения эффективности и снижения объема вычислений следует решить обратную задачу минимизации размерности признакового пространства. Последнее можно обеспечить путем определения признаков, наиболее информативных с точки зрения разделения эталонов состояний в пространстве решений. Для этого предлагается провести анализ информативности признаков априорного словаря на основе КОВ.

КОВ T строится на основе данных о значениях вектора \mathbf{y} , наблюдаемых при нахождении компьютерной сети в каждом из i диагностируемых состояний. При этом каждое i -е состояние формально описывается в виде матрицы $T^{(i)}$, содержащей множество значений вектора $\mathbf{y}^{(i)}$:

$$T^{(i)} = \begin{pmatrix} y_{11}^{(i)} & y_{12}^{(i)} & \dots & y_{1m_i}^{(i)} \\ y_{21}^{(i)} & y_{22}^{(i)} & \dots & y_{2m_i}^{(i)} \\ \dots & \dots & \dots & \dots \\ y_{n1}^{(i)} & y_{n2}^{(i)} & \dots & y_{nm_i}^{(i)} \end{pmatrix}, \quad (10)$$

где i – идентификатор классифицируемого состояния; n – количество признаков в априорном словаре A ; m_i – количество выполненных измерений сети, находящейся в i -м состоянии; $y_{jh}^{(i)}$ – значение j -го признака состояния $a_j \in A, j=\overline{1, n}$, полученное при h -м измерении, $h=\overline{1, m_i}$.

Анализ информативности признаков. В примененной схеме (1) для каждого из k диагностируемых состояний $s_i \in S$, где $i=\overline{1, k}$, КОВ определяет множество m_i объектов, которые на основе признаков из априорного словаря можно представить в виде вектор-столбцов вида $\mathbf{y}^{(i)T} = (y_{11}^{(i)}, y_{21}^{(i)}, \dots, y_{n1}^{(i)})$ в n -мерном признаковом пространстве [7]. Объединение таких векторов для всех k состояний образует исходную КОВ в виде матрицы, содержащей n строк и m столбцов:

$$T = \begin{pmatrix} y_{11}^1 & \dots & y_{1m_1}^1 & y_{11}^i & \dots & y_{1m_i}^i & y_{11}^k & \dots & y_{1m_k}^k \\ y_{21}^1 & \dots & y_{2m_1}^1 & y_{21}^i & \dots & y_{2m_i}^i & y_{21}^k & \dots & y_{2m_k}^k \\ \dots & \dots \\ y_{n1}^1 & \dots & y_{nm_1}^1 & y_{n1}^i & \dots & y_{nm_i}^i & y_{n1}^k & \dots & y_{nm_k}^k \end{pmatrix}, \quad (11)$$

где $m = m_1 + \dots + m_k$; m_i – количество объектов i -го класса. При этом для каждого $s_i \in S$ можно выделить соответствующую матрицу размерности $n \times m_i$.

Для анализа информативности признаков предлагается выполнить их сепарирование на основе использования статистических критериев однородности. С этой целью для каждой пары состояний (1:2; 1:3; ...; 1:k; 2:3; ...; 2:k; ...; (k-1):k) проверяется однородность множеств значений признаков из априорного словаря A – строк матрицы T . В качестве критериев однородности могут использоваться критерии Вилкоксона, Манна – Уитни и др. На основе анализа строк матрицы (11) в n -мерном множестве априорных признаков A можно выделить три подмножества $A^{(1)}, A^{(2)}, A^{(3)}$, для которых выполняется $A = A^{(1)} \cup A^{(2)} \cup A^{(3)}$, $n = n1+n2+n3$. Соответственно все признаки состояния a_j (где $j = \overline{1, n}$) могут быть разделены на три вида по следующему правилу:

- если для всех пар классов подтвердились гипотезы об однородности выборок значений j -го признака, признак a_j относится к *первому виду*;
- если для всех пар классов оказалось, что выборки значений признака a_j для двух сравниваемых классов подтвердили гипотезу об их статистической неоднородности, этот признак относится ко *второму виду*;
- если для признака a_j не выполнилось ни одно из двух предыдущих условий, он относится к признакам *третьего вида*.

Далее для построения модели состояния КС используются признаки второго вида. Если в результате сепарирования подмножество $A^{(2)}$ оказалось непустым (т. е. $n2 \neq 0$), то пространство принятия решений пригодно для достоверной классификации состояний [8]. Формируется уточненное пространство принятия решений $A^* = A^{(2)}$, а из матриц $T^{(i)}$ удаляются неиспользуемые строки. Если же подмножество $A^{(2)}$ оказалось пустым (т. е. $n2 = 0$), априорный словарь должен быть сформирован заново или требуется уточнение постановки задачи.

Результатами этапа анализа информативности признаков являются:

1. Обоснование факта возможности выполнения диагностики, т. е. существования эталонов диагностируемых состояний, которые формально представляются в виде компактных и разделенных образов в пространстве принятия решений.

2. Снижение размерности исходного множества признаков, построение пространства принятия решений A^* .

3. Формирование множества эталонов состояний $E = \{E_1, E_2, \dots, E_k\}$, которые представляют собой классифицированные группы вектор-столбцов КОВ, полученные в результате сокращения априорного словаря признаков.

4. Классификация состояния

Для комплексного анализа состояния сложных информационно-технических сетей наиболее применимы методы кластерного анализа, которые базируются на использовании метрик расстояний между матрицами-эталоном E_i и матрицей диагностируемого состояния, рассматриваемой как дополнительный эталон E_d . Выбор метрик зависит от особенностей диагностической задачи. Могут быть использованы расстояние Хэмминга, евклидово расстояние, обобщенное расстояние Махаланобиса и др. Для оценки расстояния между кластерами в зависимости от задачи могут быть использованы следующие метрики: расстояние ближайшего соседа, расстояние дальнего соседа, расстояние между центрами тяжести, обобщенное расстояние по Колмогорову и др.

Каждый эталон вида (10) соответствует бесконечному множеству компактно размещенных объектов, обладающих общими свойствами и образующих *кластер*. Если такие эталонные кластеры не имеют пересекающихся областей, компактны и значительно «отдалены» друг от друга, то возможно упрощенное описание эталонного кластера на основе применения различных статистических оценок (математического ожидания, среднеквадратичного отклонения и т. д.).

В реальности исследователь обычно имеет дело с ограниченной выборкой измерений, качество которой существенно зависит от способа и условий измерений. Поэтому при диагностике состояний КС в многомерном пространстве признаков требуется использование эталонных кластеров сложной формы, точный расчет которых на практике невозможен в связи с объективным недостатком данных. В этих условиях для решения задач диагностики сети предлагается использовать метод переменных гиперсфер [9].

Для применения метода переменных гиперсфер исходными данными являются полученные в процессе анализа контрольной КС математические описания эталонов состояний сети E_i . Совокупность всех эталонов E_i образует матрицу вида

$$E = \begin{pmatrix} y_{11}^1 & \dots & y_{1m_1}^1 & y_{11}^i & \dots & y_{1m_i}^i & y_{11}^k & \dots & y_{1m_k}^k \\ y_{21}^1 & \dots & y_{2m_1}^1 & y_{21}^i & \dots & y_{2m_i}^i & y_{21}^k & \dots & y_{2m_k}^k \\ \dots & \dots \\ y_{(n2)1}^1 & \dots & y_{(n2)m_1}^1 & y_{(n2)1}^i & \dots & y_{(n2)m_i}^i & y_{(n2)1}^k & \dots & y_{(n2)m_k}^k \end{pmatrix}, \quad (12)$$

где i – идентификатор классифицируемого состояния; $n2$ – количество признаков в рабочем пространстве принятия решений A^* ; m_i – количество выполненных измерений сети, находящейся в i -м состоянии; $y_{jh}^{(i)}$ – значение j -го признака состояния $a_j \in A$, $j = \overline{1, n}$, полученное при h -м измерении, $h = \overline{1, m_i}$.

Если строки матрицы заполнялись последовательно в процессе измерения, то каждый эталон E_i может рассматриваться как неупорядоченное в пространстве признаков множество $n2$ -мерных векторов $y_j^{(i)}$, $j = \overline{1, m_i}$, построенных в соответствии с моделью формализации (3).

Тогда каждому i -му эталону соответствует некоторая область пространства значений вектора наблюдения – кластер, все точки которого могут быть заданы по следующему алгоритму:

1. *Построение гиперсфер для пары объектов.* Выбирается произвольный объект $\mathbf{y}_1^{(i)}$ и строится множество расстояний $l_j^{(i)}$ до всех остальных объектов $\mathbf{y}_j^{(i)}$, $j \neq 1$. Находится ближайший объект $\mathbf{y}_2^{(i)}$, для которого выполняется $l^{(i)} = \min_{j=2, m_i} (l_j^{(i)})$. Строятся сферы радиуса $r^{(i)} = \frac{l^{(i)}}{2}$ с центрами в вершинах векторов $\mathbf{y}_1^{(i)}$, $\mathbf{y}_2^{(i)}$. Обозначается точка касания двух сфер $\mathbf{g}^{(i)}$ с координатами $(g_1^{(i)}, \dots, g_{(n)}^{(i)})$ и строится сфера радиуса $r^{(i)}$ с центром в $\mathbf{g}^{(i)}$. Предположим, что полученные для каждой пары соседних объектов три пересекающиеся гиперсферы описывают кластер, все элементы которого считаются относящимися к i -му классу состояний сети. В общем случае для описания такого элементарного кластера G_i можно использовать следующее представление:

$$G_i = \begin{pmatrix} \mathbf{y}_j^{(i)} & r & b \\ \mathbf{g}^{(i)} & r^{(i)} & b \\ \mathbf{y}_{(j+1)}^{(i)} & r & b \end{pmatrix}, \quad (13)$$

где $j = \overline{1, m_i}$; m_i – количество векторов $\mathbf{y}^{(i)}$ объектов в исходном эталоне E_i ; $b = 1$, если центром сферы является реальный вектор \mathbf{y} ; $b = 0$, если центром сферы является точка \mathbf{g} ; $r = r^{(j)}$ для объектов с номерами $j = 1$ и $j = m_i$ (крайние точки эталона); величина $r \in [\min(r^{(j)}, r^{(j+1)}), \max(r^{(j)}, r^{(j+1)})]$ для остальных точек эталона зависит от постановки задачи диагностики.

2. *Упорядочивание множества эталонных объектов.* Аналогично (14) строятся элементарные кластеры для остальных пар ближайших объектов эталона данного класса E_i и формируется общее описание эталонного кластера i -го класса в виде упорядоченного множества гиперсфер:

$$E_i \Rightarrow G_i = \begin{pmatrix} 1 & \mathbf{y}_1^{(i)} & r^{(1)} & b \\ 2 & \mathbf{g}^{(1)} & r^{(1)} & b \\ 3 & \mathbf{y}_2^{(i)} & r & b \\ \dots & \dots & \dots & \dots \\ 2m_i - 2 & \mathbf{g}^{(m_i-1)} & r^{(m_i-1)} & b \\ 2m_i - 1 & \mathbf{y}_{m_i}^{(i)} & r^{(m_i-1)} & b \end{pmatrix}. \quad (14)$$

3. *Классификация состояния на основе эталонных кластеров.* Исследуемое состояние сети может быть задано в виде одного или нескольких значений вектора $\mathbf{y}^{(d)}$, которые образуют эталон классифицируемого состояния E_d . Тогда для решения задачи классификации состояния необходимо рассчитать «меру близости» между каждым i -м эталонным кластером $G_i : E_i \Rightarrow G_i$ и контрольным диагностируемым кластером $G_d : E_d \Rightarrow G_d$.

5. Применение метода

Рассмотренный метод диагностики разработан для автоматизации прикладных исследований КС и основан на универсальном наборе алгоритмов моделирования трафика, верификации значимости признаков и классификации состояний. Метод ориентирован на использование в случаях, когда прямые измерения характеристик сети невозможны или недостаточно информативны и строгие математические правила классификации отсутствуют. Способ извлечения

информации об объекте на основе анализа косвенных признаков широко применяется в геологии, радиолокации, медицине и других отраслях.

Распространенные методы анализа состояния сетей, в том числе компьютерных, основаны на общих для диагностики технических систем принципах, таких как измерение стандартных характеристик и выявление их текущих отклонений от нормального состояния на основе одномерных пороговых значений и сигнатур (HP OpenView, NetIQ, SolarWinds, Fluke, Snort, Bro и др.). Многокритериальный анализ КС с использованием таких систем основан на алгоритмах последовательного анализа, предполагающих поочередный расчет параметров и их иерархическую обработку [10, 11]. Данный подход дает хороший результат для поиска неисправностей технических подсистем, но менее эффективен при анализе информационных процессов, поскольку требует априорного знания набора параметров и их допустимых значений, а результат может зависеть от очередности выполняемых операций.

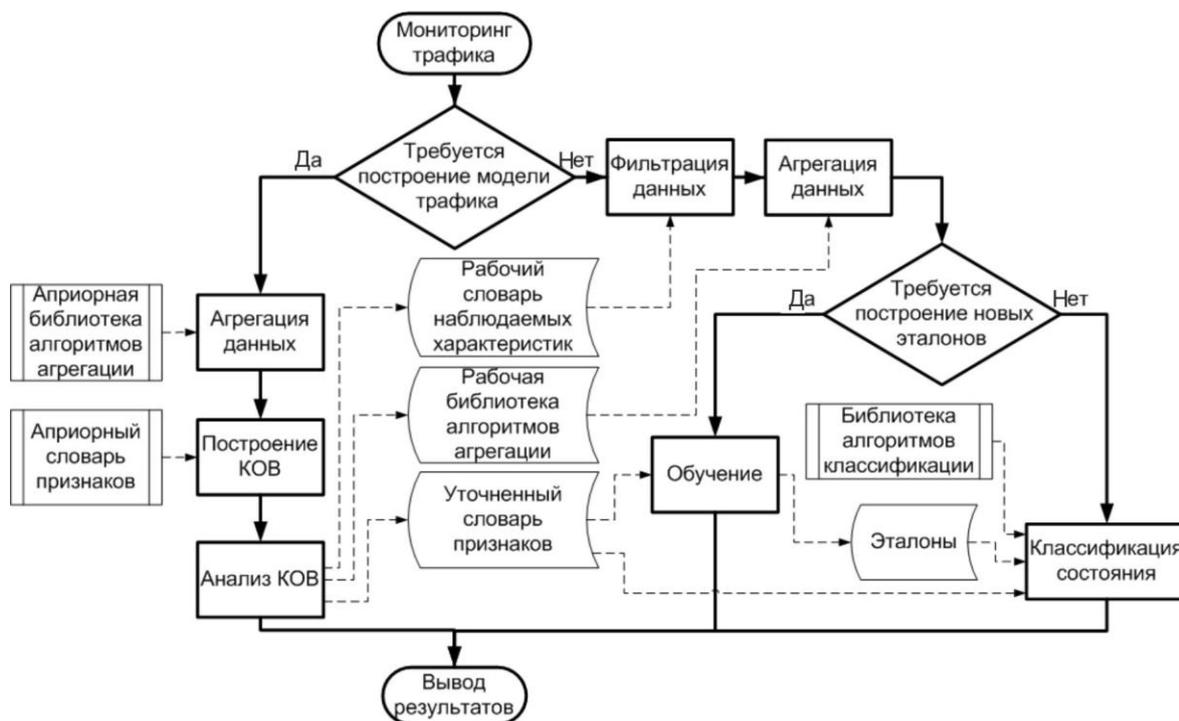
Предлагаемый авторами метод позволяет использовать результаты работы указанных и любых других диагностических средств в качестве первичных данных о трафике и на их основе выполнять не только оценку текущего состояния, но и анализ долгосрочных динамических процессов в сети. При этом набор параметров КС может формироваться и верифицироваться самой диагностической системой, а за счет применения алгоритмов кластеризации появляется возможность анализа многомерных данных, что упрощает автоматизацию процесса и позволяет получить «объемный» срез событий в КС. Таким образом, в настоящей работе под диагностикой понимается не констатация технического состояния отдельных подсистем и узлов, а комплексная оценка и прогнозирование развития сети.

Диагностическая система, основанная на предложенном методе, позволяет решать недоступные ранее классы задач, но является более сложной для эксплуатации и требует не только технической, но и математической подготовки персонала при реализации трех основных режимов работы (рисунок):

исследование – режим, позволяющий строить и верифицировать новые модели трафика, рабочие словари, библиотеки алгоритмов;

обучение – режим, позволяющий на основе существующего рабочего словаря строить базовые эталоны состояний для режима классификации сети;

классификация – режим, соответствующий режиму «диагностика» в стандартных системах.



Режимы функционирования системы диагностики

В основе работы такой диагностической системы лежат программные модули и множества данных (априорные и рабочие словари, эталоны, библиотеки алгоритмов), являющиеся в общем случае многомерными и сформированные с помощью последовательного выполнения всех режимов или готовых внешних библиотек. В статье рассмотрен один набор математических методов, но при построении диагностических систем различного назначения могут независимо изменяться первичные источники данных о трафике, алгоритмы агрегации, методы исследования однородности выборки, способы построения кластеров и оценки расстояний между кластерами.

В отличие от традиционных способов формализации рассмотренный метод диагностики использует не сами измеряемые характеристики, а их производные, что позволяет конвертировать любые типы данных в порядковый тип, пригодный для автоматизированной компьютерной обработки. В основе метода лежит предложенный способ формализации трафика (2) – (9), позволяющий учесть и использовать особенности сетевых структур, такие как множественность топологических и функциональных связей и высокую скорость протекания процессов. В качестве критериев классификации могут применяться как пороговые значения межкластерных расстояний, так и многомерные сигнатуры.

На основе знаний о многоуровневой структуре трафика КС и анализе его многомерных моделей изложенный в работе метод диагностики позволяет решать следующие прикладные задачи, требующие исследования индивидуальных свойств сети:

– поиск зависимостей в работе сегментов, устройств и сервисов сети на основе создания индивидуальных правил;

– обнаружение скрытых и аномальных явлений и процессов;

– обнаружение «медленных» изменений и построение прогнозов функционирования КС;

– оценка характера использования шифрованных соединений (р2р, VPN);

– поиск «невидимых» в сети устройств (шлюзов, неуправляемых коммутаторов, мостов);

– поиск сигнатур и сенсоров для других диагностических систем (Snort, Cisco ASA и др.).

К недостаткам метода можно отнести более низкую в общем случае достоверность результатов, обусловленную применением методов распознавания. В проведенных авторами опытах по обнаружению в сети узлов специального типа (серверов, маршрутизаторов, сканеров) получен уровень достоверности 70–90 % при применении эталонов с шестью показателями. При использовании одномерных показателей и стандартных диагностических задач система функционирует аналогично другим средствам.

Применение рассмотренного метода наиболее оправдано в следующих случаях:

– при избирательной диагностике событий, характерных и критичных для бизнес-процессов в данной сетевой системе;

– диагностике гетерогенных сетей с разнородными, в том числе неуправляемыми сетевыми устройствами;

– необходимости автоматизации принятия решений о состоянии сети.

Сложность практического применения метода состоит в его ориентации на индивидуальные свойства сети, т. е. максимальная эффективность достигается, если система диагностики реализована не в виде готового измерителя, а в виде набора настраиваемых математических инструментов, что актуально, например, для организаций, специализирующихся на аудите компьютерных систем.

Заключение

В работе предложен метод диагностики комплексных состояний компьютерной сети, основанный на апостериорном исследовании трафика КС и анализе его статистических характеристик с применением алгоритмов теории распознавания образов. Представлена реализация основных этапов построения системы диагностики компьютерной сети: формализация состояния сети, анализ информативности признаков и формирование пространства решений, построение эталонов, классификация состояний.

Разработанная модель диагностики применима для решения задач различного уровня, в том числе она может быть адаптирована к уже существующим моделям (контроль пороговых значений, поиск сигнатур и др.). В качестве источников первичных данных могут использо-

ваться любые программно-аппаратные системы журналирования сетевой активности: анализаторы сети, sniffеры и сканеры, подсистемы аудита системных событий, log-файлы серверных программ, а также результаты работы других систем диагностики.

Предложенный метод предназначен для построения автоматизированных систем диагностики, позволяющих перенести значительную часть аналитической работы на программно-технические системы, так как все этапы решения диагностической задачи формализованы и решаются с использованием методов и алгоритмов теории распознавания образов и кластерного анализа.

Перспективными направлениями развития рассмотренного метода являются: построение динамических моделей для прогнозирования работы КС на основе обработки уже полученных результатов диагностики, создание библиотек словарей признаков и эталонов для типовых задач диагностики, исследование и оптимизация интервалов агрегации для разных задач, совершенствование методов точного построения эталонов состояний.

Список литературы

1. Олизарович, Е.В. Метод и технология построения систем диагностики компьютерных сетей на основе распознавания образов : автореф. дис. ... канд. техн. наук : 05.13.13 / Е.В. Олизарович. – Гродно : ГрГУ, 2010. – 23 с.
2. Загоруйко, Н.Г. Прикладные методы анализа данных и знаний / Н.Г. Загоруйко. – Новосибирск : Изд-во Института математики СО РАН, 1999. – 268 с.
3. Айвазян, С.А. Прикладная статистика: основы моделирования и первичная обработка данных : справ. изд. / С.А. Айвазян. – М. : Финансы и статистика, 1983. – 471 с.
4. Горелик, А.Л. Методы распознавания : учеб. пособие для вузов / А.Л. Айвазян. – 3-е изд. – М. : Высш. шк., 1989. – 232 с.
5. Журавлев, Ю.И. Распознавание. Математические методы. Программная система. Практические применения / Ю.И. Журавлев. – М. : Фазис, 2005. – 159 с.
6. Олизарович, Е.В. О формализации описания компьютерной сети для диагностики на основе методов распознавания / Е.В. Олизарович, В.Г. Родченко // Информационные системы и технологии (IST'2009) : материалы V Междунар. конф.-форума, Минск, 16-17 ноября 2009 г. – Минск : А.Н. Вараксин, 2009. – С. 164–167.
7. Жукевич, А.И. Метод построения эталонов состояний компьютерной сети на основе применения алгоритмов теории распознавания образов / А.И. Жукевич, Е.В. Олизарович, В.Г. Родченко // Сетевые компьютерные технологии : сб. тр. III Междунар. науч. конф., 17–19 окт. 2007 г. – Минск : Изд. центр БГУ, 2007. – С.14–17.
8. Жукевич, А.И. Об одном методе построения компьютерной системы диагностики состояний технологических процессов / А.И. Жукевич, Е.В. Олизарович, В.Г. Родченко // VIII Междунар. конф. «Интеллектуальный анализ информации ИАИ-2008», Киев, 14-17 мая 2008 г. – Киев : Провіта, 2008. – С. 398–406.
9. Жукевич, А.И. Использование метода Монте-Карло для реализации процедуры вычисления объема пересечения сфер в пространстве R^n / А.И. Жукевич, В.Г. Родченко // Актуальные проблемы математики и компьютерного моделирования : сб. науч. ст. – Гродно : ГрГУ, 2007. – С. 192–196.
10. Бредихин, С.В. Идентификация сканеров в IP-сетях статистическим методом / С.В. Бредихин, В.И. Костин, Н.Г. Щербакова // Проблемы информатики. – 2008. – № 1. – С. 22–36.
11. Хогдал, Дж. Скотт. Анализ и диагностика компьютерных сетей / Дж. Скотт Хогдал. – М. : Лори, 2001. – 360 с.

Поступила 27.06.2012

Гродненский государственный университет им. Я. Купалы,
Гродно, ул. Ожешко, 22
e-mail: rovar@mail.ru, e.olizarovich@grsu.by, san@grsu.by

V.G. Rodchenko, E.V. Olizarovich, A.I. Zhukevich

**PATTERN RECOGNITION METHODS FOR BUILDING
A NETWORK DIAGNOSTIC SYSTEM**

A generic method for computer networks diagnosis is proposed, which is based on the theory of pattern recognition. We suggest to model the network status on the basis of the observed traffic's characteristics. Implementation mechanisms of all diagnostic stages are considered. They include formation of the dictionaries and performing procedures of training and classification.

ОБРАБОТКА ИЗОБРАЖЕНИЙ И РАСПОЗНАВАНИЕ ОБРАЗОВ

УДК 621.391.25

А.А. Борискевич, Л.А. Руис

**ГИБРИДНЫЙ АЛГОРИТМ СЖАТИЯ ИЗОБРАЖЕНИЙ БЕЗ ПОТЕРЬ
НА ОСНОВЕ ДИСКРЕТНОГО ВЕЙВЛЕТ-ПРЕОБРАЗОВАНИЯ
С АДАПТИВНЫМ ПРЕДСКАЗАНИЕМ**

Предлагается гибридный алгоритм сжатия без потерь, который основывается на использовании нового локального адаптивного дискретного лифтинг вейвлет-преобразования, обеспечивающего высокую точность аппроксимации гладких, контурных и текстурных областей изображений и минимизацию дополнительной информации об адаптации, и древовидно-блочных вейвлет-структур, позволяющих управлять соотношением быстродействие/коэффициент сжатия матрицы вейвлет-коэффициентов. Результаты моделирования предложенного алгоритма для полутоновых изображений показывают, что его эффективность сжатия без потерь превосходит алгоритм JPEG2000 на 6, 10, 8 и 8 % для тестовых изображений France, Mandrill, Barbara и Lena соответственно.

Введение

Постоянно развивающимся направлением является сжатие неподвижных полутоновых изображений. Сжатие позволяет сократить объем информации и повысить скорость ее передачи. Для систем телекоммуникаций прогрессивное сжатие изображений, основанное на алгоритмах иерархической компрессии амплитудной и позиционной информации о коэффициентах преобразования, получило наибольшее распространение. Данный подход осуществляет формирование информации о коэффициентах преобразования в порядке убывания их психовизуальной значимости. Они используются, например, в интернет-браузерах для одновременной загрузки, декодирования и ускоренного воспроизведения изображений с постепенным улучшением разрешения и качества.

К алгоритмам прогрессивного сжатия относятся JPEG 2000, EZW, SPIHT, SPECK и MECT [1–5]. Данные алгоритмы основаны на дискретном вейвлет-преобразовании (ДВП), которое обладает низкой вычислительной сложностью, свойством компактности энергии, возможностью выбора базисных вейвлет-функций с различной пространственно-частотной локализацией и гладкостью [6], и обеспечивают высокие характеристики сжатия. К ограничениям данных алгоритмов вейвлет-сжатия изображений относятся использование двухмерной древовидной или блочной вейвлет-структуры [7] и отсутствие гибкости адаптации ДВП к локальным свойствам изображений различных классов (медицинских, спутниковых, электронно-микроскопических и др.) [8].

Целью настоящей статьи является разработка алгоритма сжатия на основе адаптивного дискретного лифтинг вейвлет-преобразования (АДЛВП), обеспечивающего более высокую точность аппроксимации областей изображений с различной детальностью, минимизацию дополнительной информации об адаптации и управление соотношением быстродействие/коэффициент сжатия на основе иерархических древовидно-блочных структур.

1. Дискретное лифтинг вейвлет-преобразование на основе адаптивного предсказания в гладких и контурных областях (АДЛВП_к или ADLWT_с)

Для адаптации к локальным свойствам изображения осуществляется предварительная обработка, определяющая контурные области на основе анализа данных в окне размером 3×3 с помощью оператора Собеля [9].

Алгоритм АДЛВП_к135 [8] основан на первичном вычислении низкочастотных вейвлет-коэффициентов $s(n)$ сигнала $x(n)$ с помощью линейного оператора обновления U с весовыми коэффициентами $(1/2, 1/2)$, определяемыми базисной вейвлет-функцией Нагг с целью оптими-

зации соотношения между быстродействием, компактностью энергии и динамическим диапазоном вейвлет-изображения. Следовательно, данный оператор имеет вид

$$s(n) = U[x_o(n)] = [x_o(n) + x_e(n)]/2, \quad (1)$$

где $x_o(n)$ и $x_e(n)$ – четное и нечетное соседние значения анализируемых отчетов с индексами $n \in [1, N/2]$; $x(n) = (x_o(1), x_e(1), x_o(2), x_e(2), \dots, x_o(n), x_e(n), \dots, x_o(N/2), x_e(N/2))$.

Выбор данного подхода обусловлен тем, что только в этом случае итерации вычисления низкочастотных коэффициентов $s(n)$ линейно зависят от исходных данных на протяжении синтеза всей пирамиды вейвлет-коэффициентов и нелинейность адаптации предсказателя не влияет на $s(n)$.

Адаптация к локальным свойствам изображения позволяет использовать предсказатели $^o P$ низкого порядка O вблизи контурных линий, что повышает точность вычисления высокочастотных вейвлет-коэффициентов $d(n)$ для улучшения аппроксимации контурных линий.

Предсказатели первого, третьего и пятого порядков $O = [1, 3, 5]$ ($2/2$, $2/6$ и $2/10$ биортогональные вейвлет-функции соответственно), используемые для вычисления высокочастотных вейвлет-коэффициентов $d(n)$ в локальной окрестности, размер которой зависит от порядка предсказания O , можно представить в виде

$${}^1 P[x_o(n)] = d(n) = x_o(n) - s(n); \quad (2)$$

$${}^3 P[x_o(n)] = d(n) = x_o(n) - \left[\frac{s(n-1)}{8} + s(n) + \frac{s(n+1)}{8} \right]; \quad (3)$$

$${}^5 P[x_o(n)] = d(n) = x_o(n) - \left[\frac{-3 \cdot s(n-2)}{128} + \frac{22 \cdot s(n-1)}{128} + s(n) - \frac{22 \cdot s(n+1)}{128} + \frac{3 \cdot s(n+2)}{128} \right]. \quad (4)$$

Следует заметить, что для повышения точности вычисления вейвлет-коэффициентов на границах сигнала используется его нечетное симметричное расширение [9].

Для выбора порядка предсказания O используется функция принятия решения $D(\tilde{G}_{\text{cont}}^{x(y)}(m, n))$, характеризующая степень близости каждого нечетного пиксела к контуру в горизонтальном и вертикальном направлениях и определяемая с помощью соотношений

$$D(\tilde{G}_{\text{cont}}^{x(y)}(m, n)) = O = \begin{cases} 1 & \text{при } \tilde{G}_{\text{cont}}^{x(y)}(m, n) \geq 5, \\ 3 & \text{при } 5 > \tilde{G}_{\text{cont}}^{x(y)}(m, n) \geq 3, \\ 5 & \text{при } 3 > \tilde{G}_{\text{cont}}^{x(y)}(m, n) \geq 0; \end{cases} \quad (5)$$

$$\tilde{G}_{\text{cont}}^x(m, n) = \sum_{l=-F/2}^{(F/2)-1} H_x(l) G_{\text{cont}}^x(m, n-l); \quad (6)$$

$$\tilde{G}_{\text{cont}}^y(m, n) = \sum_{k=-F/2}^{(F/2)-1} H_y(k) G_{\text{cont}}^y(m-k, n), \quad (7)$$

где $H_x = (H_x(l) | l = -F/2, (F/2)-1) = (1, 2, 3, 4, 5, 6, 6, 5, 4, 3, 2, 1)$ и

$H_y = (H_y(k) | k = l)$ – ядра с $F=12$ весовыми коэффициентами для горизонтальной и вертикальной линейных сверток соответственно; l и k – пространственные индексы ядер

$H_x(l)$ и $H_y(k)$ соответственно; $G_{\text{cont}}^{x(y)}(m, n) = \begin{cases} 1 & \text{при } I_{\text{cont}}^{x(y)}(m, n) \geq Th \\ 0 & \text{при } I_{\text{cont}}^{x(y)}(m, n) < Th \end{cases}$ $n \in [1, N]$ – коэффициенты, характеризующие наличие или отсутствие контурной информации в горизонтальном (x) и вертикальном (y) направлениях; $I_{\text{cont}}^{x(y)}(m, n)$ – значение интенсивности пиксела с координатами (m, n) , вычисленного с помощью оператора Собеля в горизонтальном (вертикальном) направлении; $Th = \sqrt{4\mu(I_{\text{cont}}^x{}^2 + I_{\text{cont}}^y{}^2)}$ – пороговое значение для обнаружения пикселей, принадлежащих контурной области с учетом определенного коэффициента коррекции μ .

Для восстановления исходного изображения с помощью обратного АДЛВП_к с тремя (или более) разными порядками (например, $O = [1, 3, 5]$) необходимо дополнительное изображение с информацией о местоположении порядков предсказания, для кодирования которого требуется достаточно большое количество битов. Для увеличения коэффициента сжатия изображений предлагаются новые алгоритмы АДЛВП_{к13} и АДЛВП_{к15}, использующие схему с двумя порядками предсказания ($O = [1, 3]$ или $O = [1, 5]$). Они увеличивают концентрацию энергии вейвлет-коэффициентов за счет достижения оптимального соотношения между пространственным и частотным разрешениями и уменьшают количество битовых плоскостей дополнительного изображения посредством его бинаризации. В этом случае функции принятия решения $D(\tilde{G}_{\text{cont}}^{x(y)}(m, n))$ определяются с помощью соотношений

$$D_{13}(\tilde{G}_{\text{cont}}^{x(y)}(m, n)) = O = \begin{cases} 1 & \text{при } \tilde{G}_{\text{cont}}^{x(y)}(m, n) \geq 5, \\ 3 & \text{при } 5 > \tilde{G}_{\text{cont}}^{x(y)}(m, n) \geq 0; \end{cases} \quad (8)$$

$$D_{15}(\tilde{G}_{\text{cont}}^{x(y)}(m, n)) = O = \begin{cases} 1 & \text{при } \tilde{G}_{\text{cont}}^{x(y)}(m, n) \geq 3, \\ 5 & \text{при } 3 > \tilde{G}_{\text{cont}}^{x(y)}(m, n) \geq 0. \end{cases} \quad (9)$$

Для эффективного сжатия дополнительного бинарного изображения (ДБИ), несущего информацию об использовании порядков предсказания, предлагается применять стандарт JBIG (Joint Bi-level Experts Group) [10].

Для возможности управления соотношением быстродействие/коэффициент сжатия, а также для уменьшения количества значимых битов в сжатом битовом потоке предлагается древовидно-блочный алгоритм сжатия (ДБАС) матрицы вейвлет-коэффициентов (МВК), основанный на компактном представлении блочно-древовидных структур [7].

ДБАС оптимально сочетает в себе алгоритмы SPIHT и SPECK для эффективного сжатия информации о вейвлет-коэффициентах, расположенных на низкочастотных верхних и высокочастотных нижних уровнях разложения соответственно. SPECK оперирует множествами в виде квадратных блоков, состоящих из соседних вейвлет-коэффициентов внутри одного частотного поддиапазона, что позволяет быстро идентифицировать области концентрации энергии в пространственно-частотных вейвлет-областях преобразованного изображения для ограничения числа битов, требуемых для прогрессивного и компактного представления информации о субполосных вейвлет-коэффициентах. Поскольку SPECK учитывает многоуровневую кластеризацию вейвлет-коэффициентов вокруг значимых вейвлет-коэффициентов, он позволяет уменьшить корреляцию между вейвлет-коэффициентами кластерного дерева (рис. 1). В связи с этим использование алгоритма SPIHT может быть ограничено и дополнено применением алгоритма SPECK на нижних уровнях декомпозиции для уменьшения глубины нуль-деревьев с целью повышения вероятности их появления.

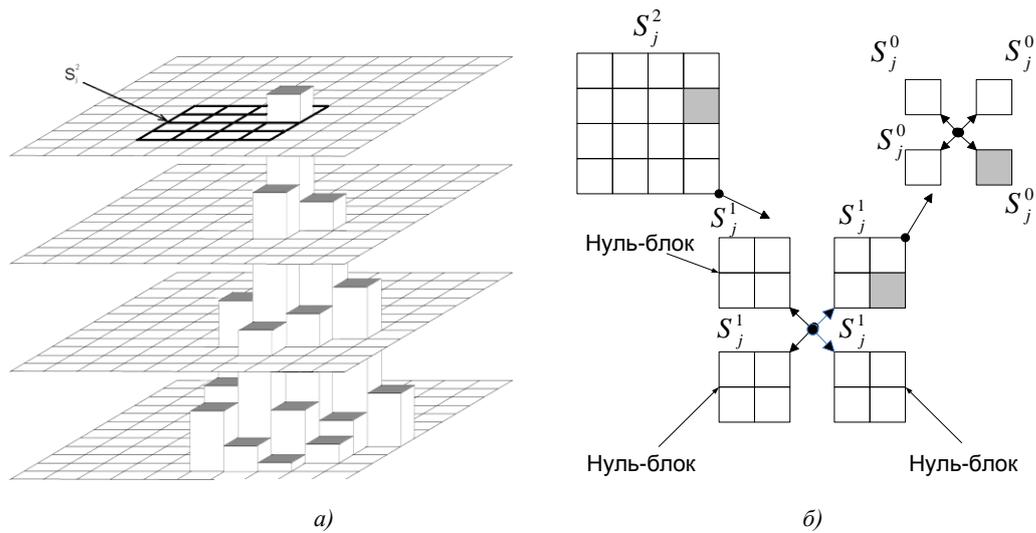


Рис. 1. Статистические особенности алгоритма сжатия SPECK: а) пространственно-частотная взаимосвязь вейвлет-коэффициентов кластерного дерева; б) уровневая кластеризация MBK на основе нуль-блоков

Процесс кластеризации вокруг значимых вейвлет-коэффициентов и формирование карт значимости для прогрессивной и эффективной передачи битов битовых плоскостей осуществляется с помощью механизма блочного деления (рис. 1). Каждый блок обозначен S_j^k , где j – уровень вейвлет-разложения исходной MBK.

В отличие от алгоритмов SPIHT и SPECK процесс кодирования расположения и знаков вейвлет-коэффициентов является составным. Он состоит из процедур кодирования расположения и знаков значимых коэффициентов на основе древовидных и блочных структур подобластей C_{SPIHT} и C_{SPECK} соответственно.

В соответствии с алгоритмом ДБАС исходная MBK разбивается на области, одна из которых представляет собой определенный уровень разрешения C_{SPIHT} , а другая – множество уровней разложения C_{SPECK} (рис. 2). Жирной линией выделена область матрицы, обрабатываемая древовидным алгоритмом SPIHT, другая область обрабатывается блочным алгоритмом SPECK. Выбор размера L области для SPECK осуществляется по критерию оптимального соотношения между быстродействием и коэффициентом сжатия.

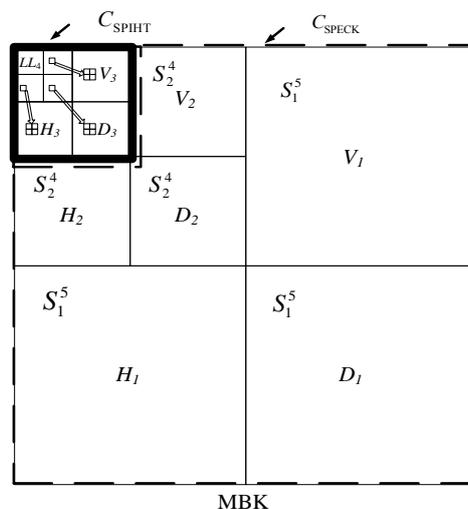


Рис. 2. Древовидно-блочное разбиение MBK с $J = 4$ уровнями разложения и $L = 2$ уровнями, обрабатываемыми SPECK

Из рис. 2 видно, что блок вейвлет-коэффициентов $S_j^k = \{H_j, V_j, D_j\}$ размером $2^k \times 2^k$ j -го уровня разложения состоит из горизонтального H_j , вертикального V_j и диагонального D_j частотных поддиапазонов при $k = \log_2(N/2^j) - 1$. Максимально возможное количество уровней разложения $J_{\max} = \lceil \log_2(N) \rceil$ зависит от размера изображения $N \times N$ при условии, что $N = 2^r$, r – целое положительное число. Размеры низкочастотного поддиапазона LL_{J+1} зависят от выбранного уровня вейвлет-разложения $J \leq J_{\max}$, удовлетворяющего требуемому соотношению между быстродействием и коэффициентом сжатия.

Древовидно-блочное разбиение МВК при $L > 1$ обеспечивает полную масштабируемость по качеству Q и частичную масштабируемость по разрешению R . Степень масштабируемости по разрешению зависит от количества уровней разложения, обрабатываемых SPECK. Конечным результатом ДБАС является битовый поток, частично масштабируемый по разрешению R и полностью масштабируемый по качеству Q (рис. 4).

Таким образом, предложенный гибридный алгоритм сжатия полутоновых изображений основан на использовании АДЛВП_к с двумя порядками предсказания, стандарта JBIG для сжатия ДБИ и ДБАС для сжатия МВК (рис. 3).

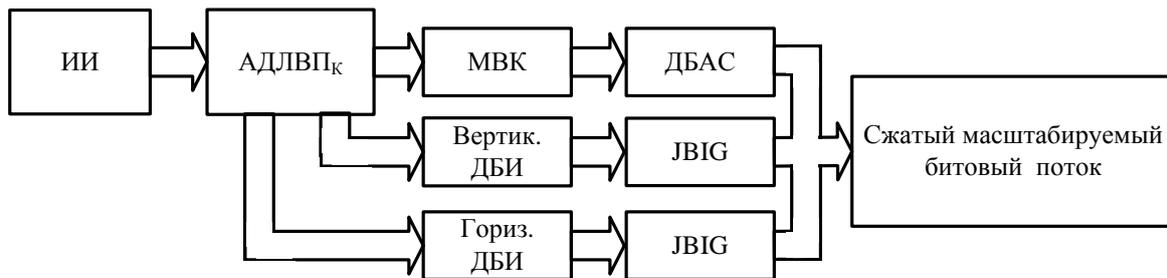


Рис. 3. Блок-схема кодера гибридного алгоритма сжатия полутоновых изображений

Структура сжатого битового потока (рис. 4), формируемая предложенным алгоритмом, состоит из четырех частей: Q и R (масштабируемых частей битового потока, обусловленных использованием алгоритма ДБАС), BI_v и BI_h (вертикальной и горизонтальной частей битового потока, обусловленных использованием стандарта JBIG).

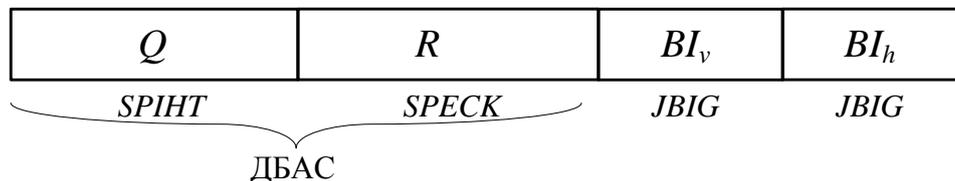


Рис. 4. Структура сжатого масштабируемого битового потока, формируемая гибридным алгоритмом сжатия без потерь полутоновых изображений

2. Результаты моделирования

Для оценки эффективности предложенного алгоритма были использованы тестовые изображения Lena, Barbara, Mandrill и France размером 512×512 (рис. 5), вейвлет-функции

Naar, 5/3, 7/5, Бинарная и 9/7 без адаптации и вейвлет-функции АДЛВПГ [11] (алгоритма вычисления адаптивного обобщенного лифтинг-вейвлет-преобразования без дополнительной информации) [11]. Проведено сравнение алгоритмов кодирования на основе блочно-древовидных структур, SPIHT и SPECK. Для вычислений были использованы программное обеспечение MATLAB и компьютер с двухъядерным процессором 2,4 ГГц и оперативной памятью 3 Гб.

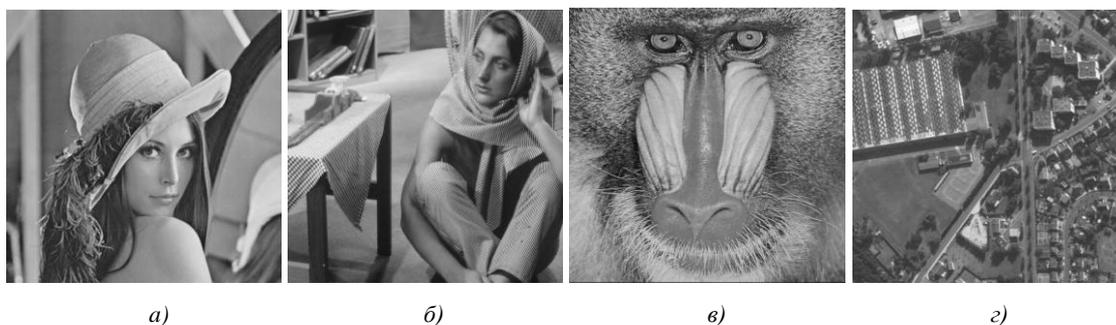


Рис. 5. Тестовые изображения размером 512x512: а) Lena; б) Barbara; в) Mandrill; г) France

Сравнительный анализ эффективности предложенного алгоритма по критерию компактности энергии вейвлет-функций представлен в табл. 1.

Таблица 1

Количество нулевых вейвлет-коэффициентов для вейвлет-функций с адаптацией и без адаптации

Изображение	Уровень разложения	Вейвлет-функции						
		Naar	5/3	7/5	9/7	АДЛВП _к 135	АДЛВП _к 13	АДЛВП _к 15
France	1	12806	17213	17621	40328	70352	47105	71026
	2	15727	19849	19732	44379	85631	57054	86506
	3	16282	20294	19931	44777	88465	58985	89368
Mandrill	1	6589	8626	8941	20071	36281	26232	36338
	2	8289	10163	10165	22322	45814	33282	45967
	3	8747	10477	10361	22653	48339	35072	48529
Barbara	1	15734	21324	21603	50862	84897	56410	85326
	2	20042	25763	24992	58165	106413	70866	107250
	3	21077	26665	25510	59087	111457	73823	112452
Lena	1	21482	26617	26892	58842	98456	76302	98970
	2	27382	32126	31169	67833	126434	95954	127273
	3	28676	33325	31859	69057	132623	100139	133619

Из табл. 1 видно, что эффективность использования вейвлет-функций алгоритма АДЛВП_к15 по критерию компактности энергии (количество нулевых вейвлет-коэффициентов) превосходит вейвлет-функции Naar, 5/3, 7/5 и 9/7 без адаптации приблизительно в 5, 4, 4 и 2 раза соответственно для всех уровней вейвлет-разложения. Определено, что количество нулевых вейвлет-коэффициентов увеличивается до третьего уровня разложения для всех вейвлет-функций и остается постоянным при более высоких уровнях декомпозиции.

Результаты оценки влияния количества уровней разложения, обрабатываемых алгоритмом SPECK в древовидно-блочном алгоритме сжатия, на выбор оптимального соотношения между скоростью и эффективностью сжатия приведены в табл. 2.

Таблица 2

Оценка влияния размера L области C_{SPECK} на характеристики сжатия ДБАС

Изображение	Алгоритм сжатия	АДЛВП _к 15		
		Кс	Степень сжатия, %	Время сжатия, с
France	ДБАС-1	1,9372	51,62	1,51321
	ДБАС-2	1,9450	51,41	1,56001
	ДБАС-3	1,9483	51,33	1,74721
Mandrill	ДБАС-1	1,5972	62,61	1,57561
	ДБАС-2	1,6079	62,19	1,60681
	ДБАС-3	1,6109	62,08	1,79401
Barbara	ДБАС-1	2,0516	48,74	1,52881
	ДБАС-2	2,0668	48,38	1,54561
	ДБАС-3	2,0683	48,35	1,57441
Lena	ДБАС-1	2,3203	43,10	1,47561
	ДБАС-2	2,3285	42,95	1,49761
	ДБАС-3	2,3309	42,90	1,52881

Из табл. 2 видно, что эффективность использования вейвлет-функций алгоритмов АДЛВП_к15 и ДБАС-3 (три уровня SPECK) по коэффициенту сжатия (Кс) и степени сжатия (отношению объема сжатой МВК к объему исходной) более чем на 0,3 % превосходит ДБАС-1 и менее чем на 0,08 % ДБАС-2 для максимального количества уровней вейвлет-разложения $J_{\text{max}} = 9$ тестовых изображений. Отметим, что при увеличении количества уровней разложения для обработки SPECK незначительно увеличивается коэффициент сжатия МВК. Эффективность ДБАС-1 по критерию быстродействия превосходит ДБАС-3 приблизительно на 200 мс и больше, а ДБАС-2 – на 190 мс и меньше для максимального количества уровней вейвлет-разложения $J = 9$ тестовых изображений. Отметим, что при увеличении L значительно уменьшается быстродействие ДБАС.

Результаты моделирования показывают, что наилучшим из алгоритмов ДБАС является ДБАС-2, так как он обеспечивает оптимальный компромисс между быстродействием и коэффициентами сжатия за счет оптимального выбора размера L области C_{SPECK} .

Сравнительные оценки коэффициентов сжатия без потерь для различных вейвлет-функций при использовании алгоритмов сжатия SPIHT, SPECK и ДБАС-2 приведены в табл. 3, а для вейвлет-функций алгоритмов АДЛВП_к13, АДЛВП_к15 при использовании SPIHT, SPECK, ДБАС-2 и JPEG 2000 – в табл. 4.

Таблица 3

Коэффициенты сжатия без потерь для различных вейвлет-функций при использовании SPIHT, SPECK и ДБАС-2

Изображение	Алгоритм сжатия	Вейвлет-функции							
		Haar	5/3	9/7	Бинарная	7/5	АДЛВП _к 135	АДЛВП _к 15	АЛВПГ
France	SPIHT	1,405	1,531	1,542	1,016	1,521	1,903	1,917	1,530
	SPECK	1,423	1,539	1,544	1,026	1,522	1,935	1,949	1,536
	ДБАС-2	1,421	1,536	1,544	1,025	1,522	1,931	1,945	1,535
Mandrill	SPIHT	1,217	1,281	1,292	1,000	1,287	1,561	1,564	1,266
	SPECK	1,247	1,294	1,295	1,000	1,291	1,608	1,612	1,275
	ДБАС-2	1,245	1,292	1,294	1,000	1,291	1,604	1,608	1,274
Barbara	SPIHT	1,444	1,578	1,609	1,049	1,566	1,994	2,007	1,556
	SPECK	1,475	1,602	1,622	1,058	1,578	2,054	2,069	1,570
	ДБАС-2	1,474	1,601	1,622	1,057	1,578	2,052	2,067	1,569
Lena	SPIHT	1,621	1,757	1,784	1,120	1,747	2,243	2,260	1,749
	SPECK	1,656	1,781	1,790	1,134	1,753	2,313	2,332	1,771
	ДБАС-2	1,655	1,779	1,790	1,133	1,754	2,310	2,328	1,770

Таблица 4

Коэффициенты сжатия полутоновых изображений для предложенного гибридного алгоритма на основе АДЛВП_к13, АДЛВП_к15 и JPEG2000

Изображение	Алгоритм сжатия	Вейвлет-функции		JPEG 2000
		АДЛВП _к 13	АДЛВП _к 15	
France	SPINT	1,5314	1,7621	
	SPECK	1,5505	1,7892	1,5902
	ДБАС-2	1,5479	1,7855	
Mandrill	SPINT	1,3601	1,4680	
	SPECK	1,3960	1,5096	1,3092
	ДБАС-2	1,3940	1,5065	
Barbara	SPINT	1,6327	1,8806	
	SPECK	1,6645	1,9344	1,6713
	ДБАС-2	1,6635	1,9328	
Lena	SPINT	1,8844	2,1013	
	SPECK	1,9157	2,1635	1,8529
	ДБАС-2	1,9145	2,1607	

Из табл. 3 видно, что эффективность вейвлет-функций алгоритма АДЛВП_к15 по критерию коэффициента сжатия превосходит эффективность вейвлет-функций Naar, 5/3, 7/5, бинарную, 9/7 без адаптации и вейвлет-функций алгоритмов АДЛВП_к13, АЛВПГ для максимального количества уровней вейвлет-разложения $J_{\max} = 9$ тестовых изображений. Алгоритм ДБАС-2 является более быстрым, чем SPECK, при сохранении коэффициента сжатия. Из результатов моделирования (табл. 4) следует, что эффективность гибридного алгоритма с АДЛВП_к15 превосходит эффективность алгоритма JPEG2000 относительно коэффициента сжатия без потерь на 6, 10, 8 и 8 % для тестовых изображений France, Mandrill, Barbara и Lena соответственно.

Заключение

В статье предложен гибридный алгоритм сжатия без потерь, основанный на новом локальном АДЛВП для повышения точности аппроксимации областей изображений с различной детализацией, на минимизации информации о бинарных изображениях и на иерархических вейвлет-структурах, позволяющий управлять соотношением между быстродействием и коэффициентом сжатия МВК.

Разработанный алгоритм АДЛВП_к15 обеспечивает увеличение концентрации энергии изображения в минимальном количестве вейвлет-коэффициентов в 5, 4, 4 и 2 раза по сравнению со стандартными вейвлет-функциями за счет выбора оптимального соотношения между пространственным и частотным разрешениями и восстановление изображения без потерь.

Установлено, что значимые биты на четвертой и выше битовых плоскостях составляют 11,25 % (SPINT), 11,37 % (SPECK), 11,32 % (ДБАС) от общего количества значимых битов (или 8,44 % (SPINT), 8,51 % (SPECK), 8,47 % (ДБАС) от общего числа всех битов) сжатого битового потока. Число значимых битов в SPINT значительно больше, чем у SPECK и ДБАС, что свидетельствует о более высокой помехоустойчивости последних. Учет соотношения количества значимых битов для различных битовых плоскостей позволит более эффективно осуществить помехоустойчивое кодирование сжатого битового потока по критерию минимизации избыточности, искусственно вносимой в передаваемый поток.

Определено, что эффективность предложенного алгоритма ДБАС-2 с АДЛВП_к15 превосходит алгоритм JPEG2000 в понятии коэффициента сжатия без потерь на 6, 10, 8 и 8 % для тестовых изображений France, Mandrill, Barbara и Lena соответственно.

Список литературы

1. Конопелько, В.К. Многомерные технологии сжатия, защиты и коммутации изображений / В.К. Конопелько, А.А. Борискевич, В.Ю. Цветков. – Минск : Белпринт, 2008. – 162 с.

2. Борискевич, А.А. Метод масштабируемого вложенного кодирования изображений на основе иерархической кластеризации вейвлет-структур / А.А. Борискевич, В.Ю. Цветков // Доклады НАН Беларуси. – 2009. – Т. 53, № 3. – С. 43–53.
3. Shapiro, J.M. Embedded Image Coding Using Zerotrees Of Wavelet Coefficients / J.M. Shapiro // IEEE Transactions on Signal Processing. – 1993. – No. 12 (41). – P. 3445–3462.
4. Said, A. New fast and efficient image codec based on Set Partitioning In Hierarchical Trees / A. Said, W.A. Pearlman // IEEE Transactions on Circuits and Systems for Video Technology. – 1996. – № 3 (6). – P. 243–250.
5. Pearlman, W.A. Efficient, low-complexity image coding with a set partitioning embedded block coder / W.A. Pearlman [et al.] // IEEE Trans. Circuits and Systems for Video Technology. – 2004. – Vol. 14. – P. 1219–1235.
6. Sweldens, W. The lifting scheme: A new philosophy in biorthogonal wavelet constructions / W. Sweldens // Proc. of SPIE. – 1995. – Vol. 2569. – P. 68–79.
7. Борискевич, А.А. Алгоритм вейвлет-сжатия полутоновых изображений на основе блочных и древовидных структур / А.А. Борискевич, О.В. Курьян // Материалы XIII Междунар. науч.-техн. конф. «Совр. средства связи». – Минск, 2008. – С. 38.
8. Борискевич, А.А. Дискретное лифтинг вейвлет-преобразование с первичным обновлением и адаптивным предсказанием на основе локального критерия гладкости изображения / А.А. Борискевич, Л.А. Руис // Известия НАН Беларуси. Сер. физ.-техн. наук. – 2011. – № 4. – С. 102–109.
9. Гонсалес, Р. Цифровая обработка изображений / Р. Гонсалес, Р. Вудс. – М. : Техносфера, 2005. – 1072 с.
10. Сэломон, Д. Сжатие данных, изображений и звука / Д. Сэломон. – М. : Техносфера, 2004. – 368 с.
11. Руис, Л.А. Алгоритм вычисления адаптивного обобщенного лифтинг вейвлет-преобразования без дополнительной информации / Л.А. Руис, А.А. Борискевич // Телекоммуникации: сети и технологии, алгебраическое кодирование и безопасность данных : материалы Междунар. науч.-техн. семинара. – Минск : БГУИР, 2011. – С. 26–32.

Поступила 14.11.12

*Белорусский государственный университет
информатики и радиоэлектроники,
Минск, П. Бровки, 6
e-mail: anbor@bsuir.by*

A.A. Boriskevich, L.A. Ruiz

HYBRID NO-LOSS IMAGE COMPRESSION ALGORITHM BASED ON DISCRETE WAVELET TRANSFORMATION WITH ADAPTIVE PREDICTION

A hybrid no-loss compression algorithm is proposed. It is based on a new local adaptive lifting discrete wavelet transformation, which provides higher accurate approximation of smooth, edge and texture image areas and minimization of side information on adaptation, and on a tree-block wavelet structures that control the trade-off between the speed and compression ratios. The simulation results of the proposed algorithm show that its compression efficiency is superior to standard JPEG2000 algorithm for test gray scale images France, Mandrill, Barbara and Lena on 6, 10, 8 and 8 %, respectively.

УДК 004.93'1; 004.932

Д.В. Прадун

СНИЖЕНИЕ РАЗМЕРНОСТИ ОБУЧАЮЩИХ ВЫБОРОК ПРИ РАСПОЗНАВАНИИ ОБРАЗОВ НА КОСМИЧЕСКИХ ИЗОБРАЖЕНИЯХ С ПОМОЩЬЮ МЕТОДА ГЛАВНЫХ КОМПОНЕНТ

Описываются сущность метода главных компонент и задача снижения размерности в процессе статистической обработки. Приводится способ вычисления главных компонент на основе определения собственных значений ковариационной матрицы. Описываются алгоритмы практической реализации метода главных компонент на основе QR-алгоритма. Проводится анализ возможности использования метода главных компонент при классификации космических изображений с целью снижения размерности обучающих выборок.

Введение

Методы классификации космических изображений с использованием обучающих выборок наряду с классификацией по эталону представляют собой наиболее распространенные способы группировки исследуемых объектов в классы. Процедура обучения играет важную роль при реализации алгоритма распознавания образов, так как от качества ее выполнения зависит корректность отнесения образов к тому или иному классу. Для этого обучающая выборка должна содержать достаточную для классификации информацию о классах и эталонах, к которым те или иные анализируемые образы относятся. Данное требование не всегда может быть выполнено, так как имеющиеся базы обучающих выборок часто содержат «избыточную» информацию о признаках анализируемых объектов, которая при использовании в процессе обучения ухудшает качество классификации и увеличивает время обработки.

Метод главных компонент (МГК) как способ снижения размерности многомерных данных при классификации статистических данных применяется в мировой практике довольно широко на протяжении нескольких лет. Примерами использования данного метода являются разработки по распознаванию лиц на изображениях [1–3], исследования в области спектроскопии [4], охраны окружающей среды [5] и многих других. Привлекательность МГК в тех или иных областях исследований обусловлена относительной простотой его использования, а также хорошей теоретической и практической проработкой при реализации метода в соответствующих системах анализа статистических данных.

В статье приведено общее техническое решение возможности использования МГК как процедуры снижения размерности (объема) обучающих выборок при классификации с обучением. Данное исследование было сделано в первую очередь для проверки влияния МГК на качество и быстродействие реализации классификаторов с обучением, а также возможности параллельной модификации метода.

1. Сущность метода главных компонент

МГК представляет собой один из способов снижения размерности анализируемых многомерных признаков, широко используемый в статистике [6, 7]. Наряду с факторным анализом и многомерным шкалированием, представляющими собой альтернативные методы снижения размерности, МГК нашел широкое применение именно при решении прикладных статистических задач [8], а также реализован в некоторых программных пакетах. Сущность МГК состоит в следующем.

Пусть имеются многомерные статистические данные вида

$$X_i = \begin{pmatrix} x_i^{(1)} \\ x_i^{(2)} \\ \dots \\ x_i^{(p)} \end{pmatrix}, \quad i = 1, 2, \dots, n. \quad (1)$$

Если число p слишком велико, возникает задача представления каждого наблюдения из (1) в виде вектора Z некоторых вспомогательных показателей $z^{(1)}, z^{(2)}, \dots, z^{(p')}$ со значением $p' \ll p$. Другими словами, при определенной p' -мерной вектор-функции $Z = Z(X)$ исходных переменных $x^{(1)}, x^{(2)}, \dots, x^{(p)}$ задается мера информативности $I_{p'}(Z(X))$ p' -мерной системы признаков $Z = (z^{(1)}(X), z^{(2)}(X), \dots, z^{(p')}(X))$ [6, 7].

Задача МГК, как и других методов снижения размерности, заключается в определении такого набора признаков \tilde{Z} , найденного в классе F допустимых преобразований исходных показателей $x^{(1)}, x^{(2)}, \dots, x^{(p)}$, что

$$I_{p'}(\tilde{Z}(X)) = \max_{Z \in F} \{I_{p'}(Z(X))\}. \quad (2)$$

Для этого определим в качестве класса $F(X)$ все возможные линейные ортогональные нормированные комбинации, такие, что

$$\begin{aligned} z^{(j)}(X) &= c_{j1}(x^{(1)} - \mu^{(1)}) + \dots + c_{jp}(x^{(p)} - \mu^{(p)}); \\ \sum_{v=1}^p c_{jv}^2 &= 1, \quad j = 1, 2, \dots, p'; \\ \sum_{v=1}^p c_{jv}c_{kv} &= 0, \quad j, k = 1, 2, \dots, p', \quad j \neq k, \end{aligned} \quad (3)$$

где $\mu^{(v)}$ – математическое ожидание $x^{(v)}$; c_{jv} – произвольные постоянные числа [6]. Мера информативности p' -мерной системы признаков $(z^{(1)}(X), z^{(2)}(X), \dots, z^{(p')}(X))$ в этом случае определяется как

$$I_{p'}(Z(X)) = \frac{Dz^{(1)} + \dots + Dz^{(p')}}{Dx^{(1)} + \dots + Dx^{(p)}}, \quad (4)$$

где D – дисперсия соответствующей случайной величины [6, 7].

При таком определении класса преобразований $F(X)$ для любых фиксированных $p' = 1, 2, \dots, p$ вектор вспомогательных переменных $\tilde{Z}(X) = (\tilde{z}^{(1)}(X), \tilde{z}^{(2)}(X), \dots, \tilde{z}^{(p')}(X))$ определяется как линейная комбинация $\tilde{Z} = LX$, где

$$L = \begin{pmatrix} l_{11} & \dots & l_{1p} \\ \dots & \dots & \dots \\ l_{p'1} & \dots & l_{p'p} \end{pmatrix}, \quad (5)$$

а ее строки удовлетворяют условию ортогональности. При этом

$$I_{p'}(\tilde{z}^{(1)}(X), \dots, \tilde{z}^{(p')}(X)) = \max_{Z(X) \in F} I_{p'}(Z(X)).$$

Полученные таким образом переменные $\tilde{z}^{(1)}(X), \tilde{z}^{(2)}(X), \dots, \tilde{z}^{(p')}(X)$ являются главными компонентами вектора X . Таким образом, *первой главной компонентой* $\tilde{z}^{(1)}(X)$ исследуемой системы показателей $X = (x^{(1)}, x^{(2)}, \dots, x^{(p)})'$ называется такая нормированно-центрированная линейная комбинация этих показателей, которая среди всех прочих нормированно-центрированных линейных комбинаций переменных $x^{(1)}, x^{(2)}, \dots, x^{(p)}$ обладает наибольшей дисперсией. Тогда *k-я главная компонента* ($k = 2, 3, \dots, p$) исследуемой системы показателей $X = (x^{(1)}, x^{(2)}, \dots, x^{(p)})'$ представляет собой такую нормированно-центрированную линейную комбинацию этих показателей, которая не коррелирована с предыдущими ($k - 1$) главными компонентами и среди всех прочих нормированно-центрированных и не коррелированных с предыдущими ($k - 1$) главными компонентами линейных комбинаций переменных $x^{(1)}, x^{(2)}, \dots, x^{(p)}$ обладает наибольшей дисперсией [6, 7].

Поскольку решение задачи, т. е. вид матрицы линейного преобразования L , зависит только от элементов ковариационной матрицы Σ , которые, в свою очередь, не изменяются при замене исходных переменных $x^{(j)}$ переменными $x^{(j)} - c^{(j)}$, исходную систему показателей необходимо центрировать, т. е. $E x^{(j)} = 0, j = 1, 2, \dots, p$. В статистической практике этого добиваются переходя к наблюдениям $\tilde{x}_i^{(j)} = x_i^{(j)} - \bar{x}^{(j)}$, где $\bar{x}^{(j)} = \sum_{i=1}^n x_i^{(j)} / n$.

Из определения главных компонент следует, что для вычисления первой главной компоненты необходимо решить оптимизационную задачу вида

$$\begin{cases} D(l_1 X) \rightarrow \max; \\ l_1 l_1' = 1, \end{cases} \quad (6)$$

где l_1 – первая строка матрицы L из формулы (5). Учитывая центрированность переменной X ($EX = 0$) и то, что $E(XX') = \Sigma$, имеем $D(l_1 X) = E(l_1 X)^2 = E(l_1 X X l_1') = l_1 \Sigma l_1'$.

Следовательно, задача (6) может быть записана следующим образом:

$$\begin{cases} l_1 \Sigma l_1' \rightarrow \max; \\ l_1 l_1' = 1. \end{cases} \quad (7)$$

Решая первое уравнение системы через функцию Лагранжа, получаем систему уравнений для определения l_1 :

$$(\Sigma - \lambda I) l_1' = 0. \quad (8)$$

Для того чтобы существовало ненулевое решение системы (8) (а оно должно быть ненулевым, так как $l_1 l_1' = 1$), матрица $\Sigma - \lambda I$ должна быть вырожденной, т. е.

$$|\Sigma - \lambda I| = 0. \quad (9)$$

Этого добиваются за счет подбора соответствующего значения λ . Уравнение (9) относительно λ называется характеристическим для матрицы Σ . Известно, что при симметричности

и неотрицательной определенности матрицы Σ , каковой она и является как всякая ковариационная матрица, это уравнение имеет p вещественных неотрицательных корней $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$, называемых характеристическими (собственными) значениями матрицы Σ .

Учитывая, что $Dz^{(1)} = D(l_1 X) = l_1 \Sigma l_1'$ и $l_1 \Sigma l_1' = \lambda$ (последнее следует из (8) после умножения слева на l_1 с учетом $l_1 l_1' = 1$), получаем

$$Dz^{(1)}(X) = \lambda.$$

Подставляем λ_1 в систему уравнений (8) и, решая ее относительно l_{11}, \dots, l_{1p} , определяем компоненты вектора l_1 . В этом случае первая главная компонента представляет собой линейную комбинацию $z^{(1)}(X) = l_1 X$, где l_1 – собственный вектор матрицы Σ , соответствующий наибольшему собственному значению этой матрицы [6, 7].

Далее аналогично можно показать, что $z^{(k)}(X) = l_k X$, где l_k – собственный вектор матрицы Σ , соответствующий k -му по величине собственному значению l_k этой матрицы.

Таким образом, соотношения для определения всех p главных компонент вектора X могут быть представлены в виде $Z = LX$, где $Z = (z^{(1)}, \dots, z^{(p)})'$, $X = (x^{(1)}, \dots, x^{(p)})'$, а матрица L состоит из строк $l_j = (l_{j1}, \dots, l_{jp})$, $j = \overline{1, p}$, являющихся собственными векторами матрицы Σ , соответствующими собственным числам λ_j . При этом матрица L по построению является ортогональной, т. е. $LL' = L'L = 1$ [6, 7].

2. Практическая реализация метода главных компонент

Исходя из описанного выше вычисления главных компонент можно сделать вывод, что практическая реализация МГК сводится к задаче вычисления собственных значений ковариационной матрицы Σ . Решения подобной задачи стали предлагать еще в 1960-х гг., и они основывались на использовании классического QR-алгоритма [9]. Данный алгоритм позволял вычислять собственные значения заданной матрицы A не слишком высокого порядка n , не принадлежащей ни к какому специальному классу матриц и хранящейся в виде квадратного массива данных размером $n \times n$ в оперативной памяти [9].

Под QR-алгоритмом можно понимать итерационный процесс вида

$$A_k = Q_k R_k, R_k Q_k = A_{k+1}, \quad (10)$$

где $k = 1, 2, \dots$, R_k – правая (верхняя) треугольная матрица на k -й итерации при условии, что столбцы матрицы A_k берутся в порядке a_1, a_2, \dots, a_n . Если же столбцы берутся в обратном порядке, можно говорить о так называемом QL-алгоритме [10], т. е.

$$A_k = Q_k L_k, L_k Q_k = A_{k+1}.$$

Главным недостатком QR- и QL-алгоритмов была высокая трудоемкость. Чтобы ее существенно снизить, использовали следующие решения [9]:

1) приведение матрицы A к хессеберговой форме, которая сохраняется на протяжении всех QR-итераций;

2) использование сдвигов, т. е. выполнение QR-итераций для матриц вида $A_k - \mu_k I$ при подходящим образом выбранных числах μ_k ;

3) проведение двойных QR-итераций в случае, если матрица A является вещественной.

Перечисленные приемы были реализованы в QR-алгоритме, представленном в программном пакете EISPACK [6, 7].

Как уже было сказано, QR-алгоритм предназначен для вычисления собственных значений обычных квадратных матриц. Если же матрица A является разреженной, т. е. имеет большое количество нулевых элементов, то, как показали исследования [9], QR-алгоритм с какого-то момента не сможет хранить промежуточные результаты применения итераций. Поэтому были предложены альтернативы QR-алгоритму, такие как методы одновременных итераций, метод Ланцоша, метод Арнольди и др. Так, метод Ланцоша был реализован во многих программных пакетах, однако версии его программной реализации часто имели жесткие ограничения к применению и не получили дальнейшего развития [9].

С развитием вычислительных средств и появлением многопроцессорных высокопроизводительных систем появились новые возможности в реализации QR-алгоритма. В первую очередь это связано с его параллельной реализацией, которая была предложена в работах [11, 12]. Распараллеливание QR-итераций позволяет существенно ускорить процесс вычисления собственных значений матриц в случае их большого размера.

3. Использование метода главных компонент при классификации космических изображений

Как уже упоминалось, МГК является одним из эффективных способов значительного сжатия исходной информации при ее статистической обработке [6]. Применительно к процессу распознавания образов на изображениях это означает, что МГК способен сократить размеры обучающих выборок при использовании классификации с обучением с целью обеспечения классификаторов информацией, достаточной для обучения и не являющейся избыточной, а также сокращения времени ее обработки.

Для проверки данного утверждения использовался простейший классификатор на основе байесовских дискриминантных функций [13]. Применительно к классификатору была определена база обучающих выборок для класса зеленых насаждений. Кроме того, были заданы обучающие выборки для класса объектов, которые формально можно описать как «не принадлежащие зеленым насаждениям». Входные данные для МГК представляют собой некоторое число наборов критериев для классификации набора пикселей космических снимков в некоторой системе категорий («зеленые насаждения»). Число наборов критериев равняется количеству снимков, для которых предварительно были сформированы наборы пикселей, удовлетворяющих критериям. Таким образом, входные данные представляют собой несколько пар обучающих выборок (каждая пара – для отдельного снимка), содержащих набор пикселей снимка. В каждой паре один файл обучающей выборки содержит пиксели снимка, которые принадлежат зеленым насаждениям, а второй файл содержит пиксели снимка, которые не принадлежат зеленым насаждениям (рис. 1).

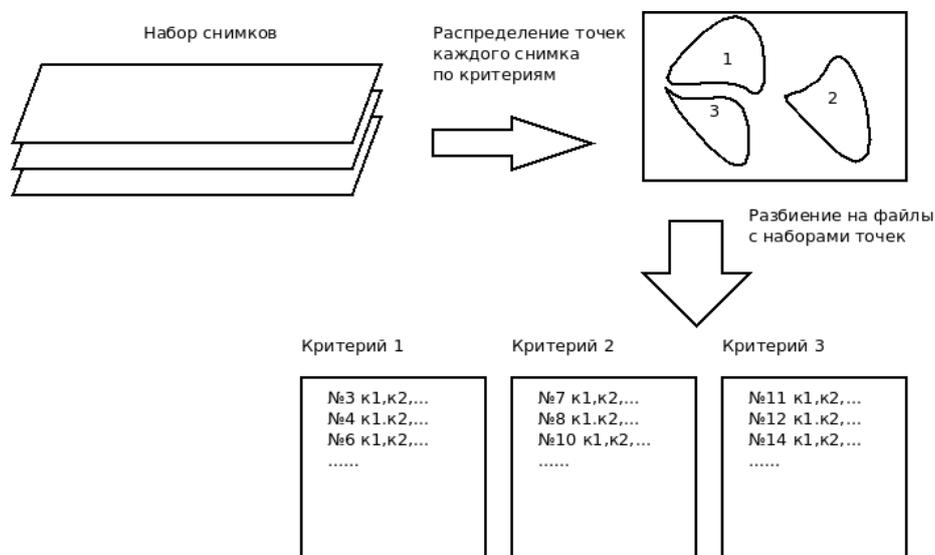


Рис. 1. Структура входных данных

Формат каждого файла обучающей выборки:

$$\langle N \rangle : \langle K1 \rangle, \dots, \langle Kn \rangle,$$

где N – индекс пиксела; $K1, \dots, Kn$ – интенсивность пиксела по каждому каналу исходного изображения.

Для программной реализации МГК использовалась программная библиотека LAPACK, которая является продолжением пакета EISPACK [14, 15]. При этом реализация МГК выполнялась в так называемом псевдопараллельном режиме: исходные данные обучающих выборок делились на блоки фиксированного размера, для каждого блока применялся МГК для расчета собственных векторов матрицы блока и затем происходило формирование усредненной матрицы по всем блокам для каждого цветового канала и критерия. Процесс разбиения на блоки и формирования усредненной матрицы использовался, во-первых, во избежание возможного переполнения оперативной памяти при обработке обучающих выборок больших размеров, а во-вторых, для проверки возможности распараллеливания программной реализации МГК.

Байесовский классификатор применялся для каждого цветового слоя космического изображения отдельно. При этом предварительно выполнялась фильтрация исходного снимка с помощью параллельного алгоритма максимального потока сети [16], а затем кластеризация отфильтрованного изображения параллельным алгоритмом нечеткой кластеризации [17]. Данная предобработка осуществлялась с целью облегчения процесса распознавания образов с точки зрения количества пикселов, исходные характеристики которых подавались на вход байесовского классификатора. Другими словами, за счет группировки пикселов в кластеры существенно снизилось количество анализируемых данных во время процесса классификации. После обработки каждого цветового слоя на выходе получается массив флагов принадлежности V каждого пиксела изображения заданным классам объектов. Если выполняется условие

$$V_{i,j} > \frac{L}{2}, \quad i=1, \dots, N, \quad j=1, \dots, L, \quad (11)$$

где L – количество цветовых слоев снимка, а N – максимальный индекс пиксела в изображении, то пиксел с индексом i определяется как принадлежащий зеленым насаждениям.

Аналогично байесовскому классификатору метод главных компонент применялся для каждого отдельного цветового слоя анализируемого снимка. В этом случае средний объем данных, подаваемый на вход МГК по обучающим выборкам каждого класса, равен K/L , где K – общий размер базы обучающих выборок для каждого класса.

В результате проведенных тестов (табл. 1) было установлено, что использование МГК позволяет значительно сократить объемы оперативной памяти, необходимой для обучения классификатора и дальнейшего распознавания образов с его помощью. Вместе с тем за счет выполнения дополнительной операции в виде процедуры снижения размерности обучающих выборок общее время обработки исходного снимка увеличивается. Поэтому выбор того, использовать МГК в процессе классификации с обучением при наличии большой базы обучающих выборок или нет, зависит от технических характеристик вычислительной системы, на которой выполняется процесс обработки. Тем не менее, если время обработки не является критическим параметром, использование МГК вполне оправдано.

Таблица 1

Показатели работы последовательного и блочно-параллельного алгоритмов

Средний объем данных обучающих выборок для цветового слоя, Мб	Время обработки*		Объем оперативной памяти, Мб	
	МГК и Байес	Байес	МГК и Байес	Байес
51,0	1 мин 16 с	45 с	24,3	35,0
518,8	7 мин 51 с	5 мин 54 с	76,4	345,6

* На вычислительной платформе Intel Core 2 Quad 2,66 ГГц, 3,24 Гб ОЗУ

В отличие от байесовского классификатора, для которого дополнительно выполнялась кластеризация данных на основе нечеткой логики, при использовании МГК в процессе классификации с помощью метода K ближайших соседей [18] время обработки исходных данных существенно сокращается (табл. 2).

Таблица 2
Показатели работы метода K ближайших соседей и МГК

Время обработки	
МГК и K ближайших соседей	K ближайших соседей
22 с 985 мс	1 мин 41 с 828 мс
7 мин 4 с	15 мин 49 с

При этом качество классификации может существенно повыситься (рис. 2). Отсюда можно сделать вывод, что эффективность использования МГК в процессе классификации с обучением повышается при выполнении классификации исходных, не обработанных предварительно графических данных. В других же случаях использование МГК зависит от технических характеристик вычислительной системы, на которой выполняется процесс обработки. При этом если время обработки не является критическим параметром, использование метода главных компонент вполне оправдано.

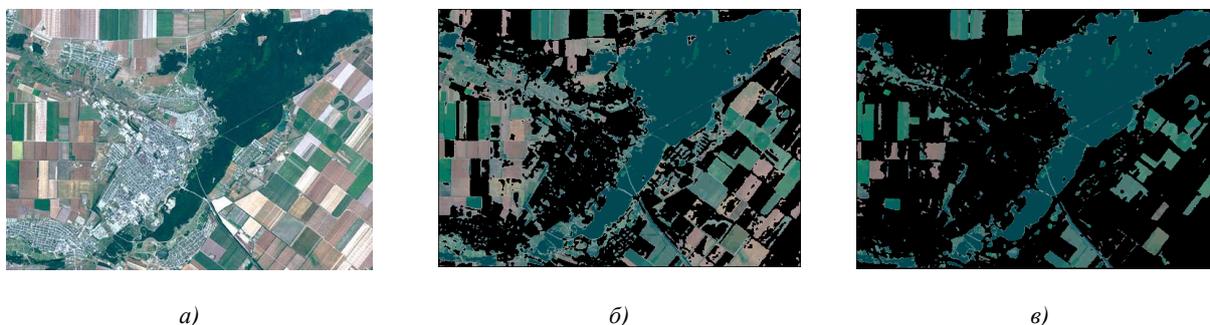


Рис. 2. Результаты классификации методом K ближайших соседей ($K = 8$): а) исходный снимок; б) классический метод; в) использование МГК для снижения размерности обучающей выборки

Заключение

МГК как один из способов снижения размерности многомерных данных позволяет существенно сократить размер базы обучающих выборок при реализации процесса классификации космических изображений. Его использование приводит к увеличению общего времени обработки данных, однако в случае когда временной фактор не является критическим, данным показателем работы МГК можно пренебречь. Кроме того, за счет разбиения входных данных обучающих выборок на блоки для каждого отдельного цветового слоя мультиспектрального изображения появляется возможность параллельной реализации МГК. При этом МГК может не только сокращать объем обучающих выборок, но и в отдельных случаях улучшать качество классификации.

Список литературы

1. Thakur, S. Face recognition using Principal Component Analysis and RBF Neural Networks / S. Thakur [et al.] // *IJSSST*. – 2009. – Vol. 10, № 5. – P. 7–15.
2. Zhang, D. Diagonal Principal Component Analysis for Face Recognition / D. Zhang, Z.-H. Zhou, S. Chen // *Pattern Recognition*. – 2006. – Vol. 39, № 1. – P. 140–142.
3. Bidyanta, N. Pattern Recognition using Principal Component Analysis / N. Bidyanta // *Binary Digits* [Electronic resource]. – 2010. – Mode of access : <https://sites.google.com/site/binarydigits10/articles/eigenface>. – Date of access : 10.09.2012.
4. Xiaoli, L. A Novel Approach to Pattern Recognition Based on PCA-ANN in Spectroscopy / L. Xiaoli, H. Yong // *Lecture Notes in Computer Science*. – 2006. – Vol. 4093. – P. 525–532.
5. Ferraz, A. The use of principal component analysis (PCA) for pattern recognition in *Eucalyptus grandis* wood biodegradation experiments / A. Ferraz [et al.] // *World Journal of Microbiology and Biotechnology*. – 1998. – Vol. 14, № 4. – P. 487–490.
6. Айвазян, С.А. Прикладная статистика. Классификация и снижение размерности. Справочное издание / С.А. Айвазян [и др.]. – М. : Финансы и статистика, 1989. – 608 с.
7. Айвазян, С.А. Прикладная статистика и основы эконометрики / С.А. Айвазян, В.С. Мхитарян. – М. : ЮНИТИ, 1998. – 1005 с.
8. Дронов, С.В. Многомерный статистический анализ : учебное пособие / С.В. Дронов. – Барнаул : Изд-во Алт. гос. ун-та, 2003. – 213 с.
9. Икрамов, Х.Д. Несимметричная проблема собственных значений. Численные методы / Х.Д. Икрамов. – М. : Наука, 1991. – 240 с.
10. Парлетт, Б. Симметричная проблема собственных значений. Численные методы; пер. с англ. / Б. Парлетт. – М. : Мир, 1983. – 384 с.
11. Stewart, G.W. A parallel implementation of the QR algorithm / G.W. Stewart // *CiteSeerX - Scientific Literature Digital Library and Search Engine* [Electronic resource]. – University park, USA, 1987. – Mode of access : <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.45.4400>. – Date of access : 22.08.2012.
12. Henry, G. A parallel implementation of the nonsymmetric QR algorithm for distributed memory architectures / G. Henry, D. Watkins, J. Dongarra // *SIAM Journal on Scientific Computing*. – 2002. – Vol. 24, № 1. – P. 284–311.
13. Гонсалес, Р. Цифровая обработка изображений / Р. Гонсалес, Р. Вудс. – М. : Техносфера, 2005. – 1072 с.
14. Smith, B.T. Matrix eigensystem routines – EISPACK guide / B.T. Smith [et al.] // *Lecture Notes in Computer Science*. – 1976. – Vol. 6. – 551 p.
15. Garbow, B.S. Matrix eigensystem routines – EISPACK guide extension / B.S. Garbow [et al.] // *Lecture Notes in Computer Science*. – 1977. – Vol. 51. – 343 p.
16. Прадун, Д.В. Блочно-параллельная кластеризация мультиспектральных изображений с помощью алгоритма максимального потока в сети / Д.В. Прадун, Б.А. Залесский // *Информатика*. – 2011. – № 2(30). – С. 12–20.
17. Прадун, Д.В. Блочно-параллельная кластеризация изображений на основе нечеткой логики / Д.В. Прадун, А.А. Кравцов // *Пятый Белорусский космический конгресс : материалы конгресса. В 2 т. (25-27 октября 2011 года, Минск)*. – Минск : ОИПИ НАН Беларуси, 2011. – Т. 2. – С. 47–53.
18. Hastie, T. The Elements of Statistical Learning. Data mining, Inference, and Prediction (Second Ed.) / T. Hastie, R. Tibshirani, J. Friedman // *Trevor Hastie – Publications* [Electronic resource]. – 2009. – Mode of access : http://www.stanford.edu/~hastie/local.ftp/Springer/ESLII_print5.pdf. – Date of access : 12.03.2012.

Поступила 23.08.12

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

D.V. Pradun

**REDUCTION OF TRAINING SAMPLES DIMENSION IN PATTERN
RECOGNITION OF SPACE IMAGES USING PRINCIPAL
COMPONENTS ANALYSIS**

The essence of principal components analysis and the problem of dimension reduction are described. A method of principal components calculation is presented, which is based on the covariance matrix eigenvalues determination. Practical implementations of principal components analysis are described, which are based on QR-algorithm. Application of principal components analysis in space images classification for the reduction of training samples dimension is discussed.

УДК 004.056.5

С.А. Сейеди, Р.Х. Садыхов

СРАВНЕНИЕ МЕТОДОВ СТЕГАНОГРАФИИ В ИЗОБРАЖЕНИЯХ

Стеганография – это метод сокрытия информации в объектах различных форматов (контейнерах). Существует большое разнообразие методов для конкретных контейнеров, в которые скрытно записывается информация. Наиболее часто в качестве контейнера используются цифровые изображения, которые без искажений могут пересылаться по компьютерным сетям. Исследуется основная методика стеганографии, ее сильные и слабые стороны. Дается сравнительная оценка основных алгоритмов стеганографии.

Введение

Назначение компьютерной безопасности состоит в защите информации от несанкционированного доступа, случайного или целенаправленного искажения данных без изменения основных свойств файлов. Криптография создавалась как методика для защиты систем связи методами кодирования и последующей расшифровки данных. Стеганография дополняет криптографию, скрывая сам факт наличия сообщения в передаваемом потоке данных. Стеганографию можно рассматривать как создание скрытого канала связи. Если криптография маскирует сообщение, то стеганография пытается скрыть наличие такого сообщения.

В табл. 1 дается сравнение трех подходов к передаче секретной информации. В криптографии для расшифровки сообщения используется секретный ключ. При перехвате сообщения недоброжелатель, если и не сможет прочесть его, то в большинстве случаев сможет исказить. Цифровая подпись подтверждает подлинность сообщения, она может быть удалена. Если недоброжелатель не знает метода инкапсуляции, стеганографическое сообщение нельзя удалить без значительного искажения контейнера.

Таблица 1

Сравнение секретных техник связи

Техника связи	Конфиденциальность	Целостность	Неустрашимость
Криптография	Да	Нет	Да
Цифровая сигнатура	Нет	Да	Нет
Стеганография	Да / Нет	Да / Нет	Да / Нет

1. Обзор методов стеганографии**1.1. Различные алгоритмы инкапсуляции сообщения**

Почти все форматы файлов годятся для стеганографических вставок, но формат файла-контейнера ограничивает допустимые методы стеганографии. Во всех форматах изображений существуют излишние биты, по крайней мере такие, значения которых практически не сказываются на качестве изображения. В эти биты файла изображения можно вставить скрытую информацию. Многие форматы изображений и аудиофайлов пригодны для вставки скрытой информации. На рис. 1 показаны категории файлов, для которых можно применять стеганографию.

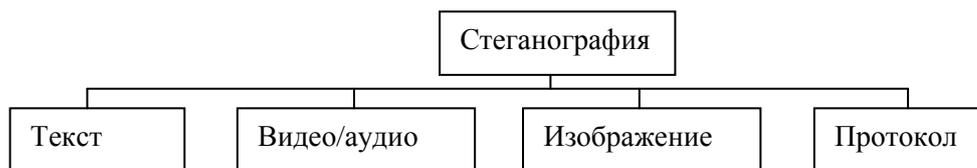


Рис.1. Различные виды приложений

1.2. Стеганографические системы

Стеганографическая система в общем виде состоит из двух компонентов: вставки и извлечения сообщения (рис. 2).

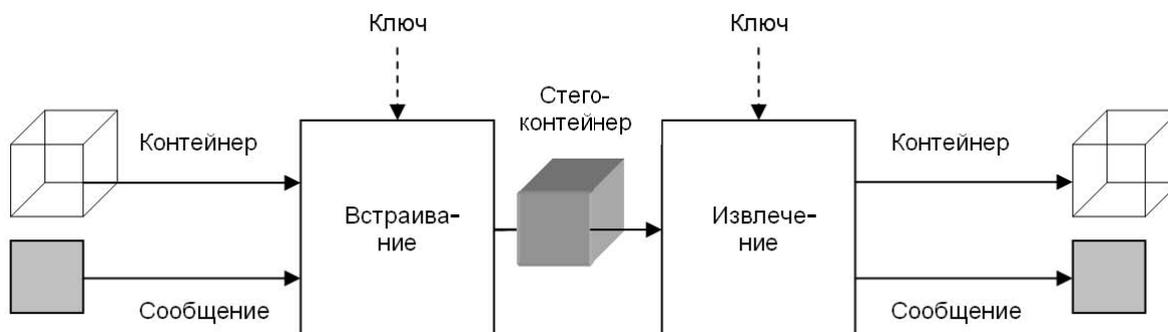


Рис. 2. Структура стеганографии

Процесс обработки стеганографического сообщения включает следующие объекты:
 сообщение – данные любого типа;
 контейнер – любая информация, пригодная для сокрытия в ней сообщений;
 стегоконтейнер – контейнер, содержащий скрытое сообщение;
 ключ (стегоключ) – секретный ключ, необходимый для шифрования (расшифровки) сообщения с целью усиления защиты.

Для обработки сообщения может использоваться пароль. Это ключевое слово, применяемое для записи и расшифровки скрытого сообщения.

2. Стеганографические изображения

Скрытое сообщение можно инкапсулировать практически во все виды данных. Большинство инструментов стеганографии ориентируется на передачу сообщения в Интернете, где значительная часть информации передается в виде изображений. При обработке изображения-контейнера учитывают формат файла, в частности методы сжатия. От этого зависят как методы инкапсуляции, так и объем стеганограммы, которую можно вставить в файл. Сложность процедур стеганографии также зависит от формата контейнера.

2.1. Классификация методов стеганографии для изображений-контейнеров

В течение всего периода развития стеганографии было разработано множество методов, зависящих от форматов изображений и от применяемого аппаратного обеспечения. Методы стеганографии для изображений можно разделить на два класса: методы для временной области [1, 2] и методы для частотной области [1–3]. Для временной области основные процедуры инкапсулируют скрытое сообщение в младшие биты цифрового кода пикселей изображения. Для частотных процедур стеганограмма вставляется в частотную характеристику изображения. Временные процедуры включают следующие методы:

- внедрение цифрового кода сообщения в изображение: незначимый или мало значимый младший бит цвета или палитры изображения заменяется битом вставляемого сообщения;
- статистические методы замены: бит изображения заменяется по некоторому статистическому закону; например биты фрагмента сообщения вставляются в псевдослучайно выбранные патчи изображения;
- частотные процедуры состоят в замене мало значимых частотных характеристик изображения, например замене некоторых коэффициентов в дискретном косинус-преобразовании (ДКП).

В частотных процедурах кроме ДКП используется дискретное вейвлет-преобразование (ДВП), эти и некоторые другие методы обычно применяются в патчах и методах расширения спектра.

3. Методы для временной области

Существуют различные методы встраивания секретного сообщения во временной области контейнера:

Метод наименее значащего бита (LSB). Во временных процедурах наиболее часто применяется метод LSB. Так, в формате градаций серого цвета с 8-битным кодированием пиксела в некоторых заданных пикселах заменяется младший бит кодировки (рис. 3).



Рис. 3. Взвешивание 8-битных пикселей

При замене некоторых младших битов пикселей изображения само изображение изменяется незначительно. Математическая формула замены имеет вид [4]

$$x_i' = x_i - x_i \bmod 2^k + m_i, \quad (1)$$

где x – i -й пиксел стегограммы; x – пиксел исходного изображения; m – десятичное значение i -го блока сообщения. Номер заменяемого бита обозначен k . Процедура раскодировки очевидна – нужно собрать последние биты измененных пикселей. Формула извлечения определяется выражением

$$m_i = x_i' \bmod 2^k. \quad (2)$$

Таким образом, перестановка извлеченных m даст исходное пересылаемое сообщение. При использовании 24-битного цветового формата RGB в стеганограмме можно использовать цветовые байты. Всего можно использовать 3 бита на один пиксел. В картинку размером 800x600 пикселей можно записать скрытое сообщение объемом около 180 КБ. Например, если представление трех пикселей такого изображения представляет собой последовательность

```
(00101101 00011100 11011100)
(10100110 11000100 00001100)
(11010010 10101101 01100011),
```

то число 200 с двоичным кодом 11001000 будет вставлено как

```
(00101101 00011101 11011100)
(10100110 11000101 00001100)
(11010010 10101100 01100011).
```

Замена потребовалась только в трех подчеркнутых из восьми битов записи. Можно заменять один или два последних бита цвета, при этом глаз человека не распознает искажения параметров изображения. Предполагая равномерность распределения числовых значений заменяемых битов, получаем, что при инкапсуляции сообщения замена в среднем потребует только в половине младших битов контейнера.

В этом методе такие параметры, как формат файла, метод сжатия изображения (предполагая сжатие без потерь) и положение бита в байте, играют важную роль. Объем сообщения и его скрытность в формате RGB лучше по сравнению с форматом градаций серого цвета. Существует

множество методов LSB инкапсуляции сообщения в формат градаций серого цвета с различными степенями успеха. Контейнер должен скрыть сообщение, LSB с одним битом применяется в форматах GIF и BMP контейнера. Файлы GIF сжимаются без потерь, они имеют небольшой размер и широко используются в Интернете. В формате GIF можно заменять младший бит в каждом байте файла. Формат BMP не использует сжатия, и файлы этого формата имеют большой объем. К сожалению, для скрытия сообщения требуется файл BMP большого объема. В наше время файлы формата BMP не передаются по Интернету и их применение может вызвать подозрение. Поэтому в стеганографии используются другие форматы.

Существуют методы LSB, заменяющие два, три и даже четыре бита цвета в формате RGB. Эти методы зависят от приложений, в которых они применяются, и здесь важно найти равновесие между качеством изображения и объемом вставляемого сообщения.

Для того чтобы затруднить выявление сообщения в стеганограмме, используются специальные методы вставки сообщения в цвета изображения, например циклическая перестановка битов сообщения, вставка только в красный байт цвета.

Цветовая палитра и LSB. В изображения, использующие заданную цветовую палитру, цвет пиксела записывают в виде индекса, обычно это 8-битное число. Например, созданный для Интернета формат GIF использует 8-битную нумерацию цветов, т. е. контейнер содержит пикселы из заданного набора 256 цветов. Для того чтобы сократить время поиска цвета по индексу в цветовой таблице, цвета упорядочивают по частоте их появления в картинке. Поэтому при однобитном методе LSB результат зависит от того, соответствует ли малое изменение индекса малому изменению цвета, неразличимому для глаз человека. Для разрешения этой проблемы существуют два способа: первый – переупорядочивание цветов по их плавному изменению, второй – добавление новых цветов, близких к табличным. Первый способ применим к изображению с градациями серого цвета [6]. В этом случае картинки практически неразличимы и выявление скрытого сообщения затруднено [5]. Для выявления скрытого сообщения и в первом и во втором случаях достаточно переупорядочить цветовую таблицу, после чего низкое качество изображения укажет на наличие вставки.

Псевдослучайные перестановки (ПП). В этом методе биты скрытого сообщения распределены по изображению случайным образом. Такой подход увеличивает трудоемкость выявления скрытого сообщения, особенно если псевдослучайный датчик генерирует последовательность сложным алгоритмом. Здесь могут возникнуть коллизии, т. е. повторение адреса заменяемого бита, что означает попытку вставки двух значений в один бит контейнера. Чтобы преодолеть коллизию, используется специальный файл, в который записываются адреса заменяемых битов. Если генерируется адрес, уже содержащийся в этом файле, то он не используется и генерируется следующее псевдослучайное число. Применение дополнительного файла усложняет как обработку стеганограммы, так и попытку выявления скрытого сообщения. Если при этом биты случайно выбираются для замены из всего поля контейнера, то такой метод может создать стеганограмму с большими и ясно различимыми повреждениями контейнера. Случай, когда старшие биты загружаются в MSB, показан на рис. 4 [6].

Контейнер	Секретное сообщение	Стегоконтейнер
(00101101 00011100 11011100)		(0 <u>1</u> 10110 <u>1</u> 000 <u>1</u> 1 <u>0</u> 00 1101 <u>0</u> 100)
(10100110 11000100 00001100)		(10 <u>1</u> 00 <u>0</u> 10 110 <u>1</u> 0100 00001100)
(11010010 10101101 01100011)ADD 11000101		(11010010 10101101 01100011)

Рис. 4. Метод псевдослучайных перестановок

Метод с использованием патчей (ИП). Этот метод является статистическим кодированием информации путем изменения некоторых статистических свойств контейнера (добавляет излишние данные к скрытому сообщению и затем размывает его по изображению с применением гауссового распределения) и использует проверку гипотез при извлечении сообщения. Секретный ключ применяется к случайно выбранному подмножеству пикселов из изображения, затем подмножество разделяется по двум патчам – А и В. Яркость одного подмножества пиксе-

лов из одного патча смещается на положительное число, а яркость пикселей из другого патча – на противоположное (отрицательное) число (рис. 5). Содержимое контейнера не зависит от изменения патчей. Контрастность в пикселях изменяется на один бит, изменение очень мало и не воспринимается визуально, тем более что средняя освещенность не изменяется [7, 8].



Рис. 5. Патч-метод с двумя областями

Метод расширения спектра (РС). В этом методе скрытое сообщение распределяется по контейнеру, что затрудняет выявление стеганограммы [9]. Метод предложен Марвелом для маскировки сообщения. Он комбинирует расширение спектра, кодирование с исправлением ошибок и обработку изображений для скрытия вставляемого сообщения [10]. Метод описывает расширение спектра с представлением сигнала в узкой полосе на широкополосное представление. Это может сопровождаться переносом и изменением профиля колебания с узкой полосы на широкую, что напоминает добавление шума. После расширения полосы передачи энергия сигнала на узкой полосе уменьшается, что еще больше затрудняет выявление скрытого сообщения [11].

Скрытое сообщение в данном методе может предварительно размываться шумом, а затем инкапсулироваться в изображение. Энергия этого сигнала будет гораздо ниже энергии контейнера, что создает дополнительные трудности выявления стеганограммы, которая становится практически незаметной для глаза человека. Без наличия исходного изображения компьютерный анализ изображения с вставкой такой стеганограммы требует больших ресурсов [10].

4. Методы преобразования области определения

Некоторые методы, например расширения спектра и использования патчей, как во временной области, так и в частотной очень похожи.

Методы, использующие частотное представление: быстрое преобразование Фурье, ДКП, ДВП [12], вначале преобразуют изображение, получая частотные коэффициенты или функцию плотности и применяя подходящий математический подход. Затем в частотную характеристику вставляется скрытое сообщение. После этого проводится обратное преобразование, возвращающее сигнал во временную область. Преимущество такого подхода заключается в возможности обработки сигналов или шума. Однако такие методы имеют высокую вычислительную сложность и вследствие этого низкую скорость работы.

Чтобы дать понятие стеганографических методов такого типа, необходимо вначале рассмотреть форматы файлов, которые преобразуются в частотную область. Одним из широко используемых форматов является JPEG, который многократно сжимает изображение. Это метод сжатия с потерей данных, несущественных для визуального восприятия [13].

4.1. Дискретное косинус-преобразование

Одна из идей стеганографии состоит в том, что скрытое сообщение маскируют заменой несущественных параметров изображения, но в формате JPEG в процессе сжатия возможно изменение или даже удаление таких параметров. Алгоритм сжатия формата JPEG состоит из по-

следовательности шагов, одна часть которых изменяет данные, а другая – сохраняет их. Шаги ДКП и квантизации изменяют данные, другие шаги, например кодировка по Хаффману, сохраняют их. Сообщение следует вставлять между этими шагами.

Алгоритм сжатия состоит из шести шагов (рис. 6) [14, 15]:

- 1) разложить изображение RGB по составляющим цветам: красному, зеленому и синему;
- 2) преобразовать цвета в формат YCbCr;
- 3) разделить матрицы интенсивности каждого цвета на блоки 8x8 пикселей;
- 4) выполнить квантизацию интенсивностей каждого блока по таблице квантизации;
- 5) произвести кодировку с учетом энтропии;
- 6) выполнить обратное ДКП.



Рис. 6. Процесс ДКП изображения в формате JPEG

4.2. Дискретное вейвлет-преобразование

Вейвлет-преобразования широко применяются в обработке сигналов и сжатии изображений. Ввиду возможности анализа сигнала в выделенном интервале спектра они являются мощным инструментом для исследования особенностей поведения сигнала на различных частотных интервалах. Один из современных форматов кодирования JPEG-2000 основан на ДВП [16].

Несмотря на все положительные стороны классического JPEG с применением ДКП, а именно простоту использования, достаточно хорошее сжатие и возможность построения специализированной аппаратуры кодировки в этом формате, классический формат обладает и некоторыми недостатками. Так как исходное изображение разделяется на блоки, возможно высокое коррелирование интенсивности блоков. Этот феномен влияет на качество изображения и известен как «блоковые искажения», он особенно заметен на пикселах с низкой интенсивностью (рис. 7). Перекрывающееся ортогональное преобразование (ЛОТ) [13] частично решает эту проблему, сглаживая границы блоков. Несмотря на очевидные улучшения, вычислительная сложность не позволила в алгоритме сжатия полностью заменить ДКП на ЛОТ.



Рис. 7. Изображение «Лена»: а) оригинальное; б) реконструированное блокированием артефактов

5. Сравнение стеганографических методов

Каждый стеганографический метод обладает как сильными, так и слабыми сторонами. Пользователю важно выбрать метод, который в наибольшей степени соответствует поставленной задаче. Существует несколько критериев для сравнения и выбора наиболее подходящего пользователю метода стеганографии. Все алгоритмы стеганографии должны удовлетворять нескольким основным требованиям. Наиболее важно, чтобы алгоритм давал малозаметное изменение изображения-контейнера. Рассмотрим критерии сравнения:

- *незаметность или уровень восприятия (нез.)*. Это главное требование – стеганограмма не должна распознаваться глазом человека. Человек не должен видеть различия между исходным изображением и тем же изображением со вставленным сообщением. Выполнение данного требования зависит от объема сообщения и формата изображения [17];

- *вместимость (вмест.)*. Это требование определяет размер вставляемого сообщения, который зависит от формата контейнера [18];

- *робастность*. Вставляемое сообщение не должно быть повреждено процессами обработки и передачи, присущими данному формату. Существует два типа робастности, один – для статистической обработки, второй – для целенаправленного повреждения стеганограммы:

- *робастность для защиты от статистической атаки (РПСА)*. Статистические тесты применяются для выявления наличия стеганограммы в контейнере, это методы статистической обработки данных, которые можно применять как во временной, так и в частотной области. Многие методы стеганографии обладают сигнатурой, которую легко выявить статистическим анализом. Стеганограмма не должна оставлять в контейнере сигнатуру [17, 18];

- *робастность для защиты от целенаправленного повреждения стеганограммы (РПЦП)*. Этот вид робастности обусловлен тем, насколько скрытое сообщение зависит от контейнера. Такие простые операции, как кадрирование, вращение и др., могут повредить сообщение;

- *способность к обнаружению или скрытность (СОС)*. Этот критерий определяет успешность метода скрытия, при распознавании наличия сообщения он обуславливает сложность алгоритма распознавания [4–9];

- *вид области (ВО)*. Этот параметр указывает на область, в которой применялась стеганография – временная (В) или частотная (Ч). Во временной области методы работают быстро, в частотной гораздо медленнее, но частотная область более надежна для скрытия сообщения;

- *независимость от формата (НФ)*. Следует использовать различные форматы файлов. Если партнеры постоянно используют один формат, то это может привести к мысли о тайной переписке [17, 18].

В табл. 2 дается сравнение методов стеганографии по указанным критериям. Высокий уровень (*H*) говорит о полном соответствии требованиям, низкий (*L*) – о несоответствии, средний уровень (*M*) указывает, что требование зависит от конкретного параметра. В одних случаях может быть высокий уровень соответствия, в других – низкий. Например, скрытность алгоритма LSB с цветовой гаммой зависит от формата файла.

Таблица 2

Сравнение стеганографических методов

Алгоритмы	Критерии						
	Нез.	Вмест.	РПСА	РПЦП	СОС	ВО	НФ
LSB(BMP)	<i>H</i>	<i>H</i>	<i>L</i>	<i>L</i>	<i>L</i>	<i>L</i>	В
LSB(JPEG)	<i>H</i>	<i>H</i>	<i>L</i>	<i>L</i>	<i>H</i>	<i>L</i>	В
LSB(палитра)	<i>M</i>	<i>M</i>	<i>M</i>	<i>L</i>	<i>M</i>	<i>L</i>	В
ПП	<i>M</i>	<i>L</i>	<i>M</i>	<i>L</i>	<i>H</i>	<i>H</i>	В
ИП	<i>H</i>	<i>L</i>	<i>H</i>	<i>H</i>	<i>H</i>	<i>H</i>	В/Ч
РС	<i>H</i>	<i>L</i>	<i>H</i>	<i>M</i>	<i>H</i>	<i>H</i>	В/Ч
ДКП	<i>H</i>	<i>L</i>	<i>M</i>	<i>M</i>	<i>H</i>	<i>L</i>	Ч
ДВП	<i>H</i>	<i>H</i>	<i>M</i>	<i>M</i>	<i>H</i>	<i>L</i>	Ч

В идеальном случае стеганографический алгоритм должен удовлетворять высоким уровням всех критериев. Среди множества рассмотренных методов авторы не смогли найти такого алгоритма. Необходим взвешенный выбор стеганографического метода, который зависит от используемого пользователем приложения. Рассмотрим пригодность различных алгоритмов для форматов файлов:

LSB для BMP. Растровый формат BMP не использует сжатия, поэтому файлы этого формата имеют большой объем. К сожалению, для сокрытия сообщения в этих файлах необходим очень большой контейнер. В наше время этот формат используется в глобальной сети нечасто и его многократное применение может вызвать подозрение. Таким образом, этот формат не удовлетворяет условиям робастности. Обычно для сокрытия сообщения используется глубина 1, 4, 8, 16, 24, 48 и 64 бита на пиксел, поэтому BMP может вместить достаточно большое сообщение. Очевидно, что чем больше битов цвета заменяется, тем больше вероятность того, что невооруженный глаз заметит повреждение контейнера.

LSB для JPEG. Распространенный формат JPEG использует 8 битов на каждый цвет RGB, всего 24 бита на пиксел. JPEG может скрыть сообщение большого объема. Обычно стеганограмму в данном формате трудно распознать, это зависит от применяемого метода. Формат JPEG использует сжатие с потерями, и в процессе сжатия сообщение может быть повреждено. В таком случае JPEG не удовлетворяет критерию робастности.

LSB для цветовой палитры. Формат GIF кодирует пиксел 8 битами, изображение записывается в 256 цветах. Алгоритм LSB скрывает информацию с различными степенями успеха в зависимости от доли изменяемых бит. Необходимо искать равновесие между безопасностью и распознаваемостью. В разд. 4.2 указано, что изменение цвета в изображении зависит от структуры цветовой палитры, в этом случае тип изображения (полутоновое или цветное) играет важную роль. Распознаваемость и РПСА зависят от типа изображения.

Псевдослучайные перестановки. Метод вставляет биты сообщения с изменением порядка их появления в сообщении, что затрудняет работу по обнаружению и расшифровке сообщения. Как и в методе LSB, сообщения, биты которых вставлены случайным образом в младший бит, могут быть повреждены. Если же вставлять сообщения в другие, не последние биты, это может внести дополнительный шум в изображение. Формат изображения для этого метода не имеет большого значения.

Метод с использованием патчей. Недостаток этого метода состоит в том, что в один патч инкапсулируется только один бит. Разбиение изображения на более мелкие фрагменты позволяет вставить больше битов [7]. Преимущество этого метода состоит в том, что сообщение распределено по всему изображению, и если один из патчей будет поврежден, то это не принесет больших потерь и сообщение можно восстановить из других патчей [4]. Однако это зависит от других обстоятельств, в частности от размера сообщения. Короткое сообщение можно дублировать. Этот метод имеет большую робастность, так как скрытое сообщение может сохраниться даже при сжатии с потерями.

Метод расширения спектра. Этот метод распределяет сообщение по всему изображению. Такую стеганограмму трудно распознать. Частотная характеристика сообщения обладает гораздо меньшей энергией, чем энергия контейнера. Поэтому распознавание присутствия скрытого сообщения в контейнере очень затруднительно. Этот метод имеет большую робастность против атак.

Дискретное косинус-преобразование. Методы области преобразования (частотной области) скрывают сообщение в значительной области изображения, что делает их более робастными по сравнению с методами во временной области, включая сжатие, обрезку и некоторые алгоритмы обработки изображений. Однако существует компромисс между количеством информации и робастностью. Встраивание информации в область ДКП проще выполнить заменой коэффициентов ДКП. Коэффициенты равны нулю и замена многих нулей ненулевыми значениями будет иметь эффект коэффициента сжатия, поэтому мощность ДКП меньше, чем у LSB.

Дискретное вейвлет-преобразование. Инкапсуляция сообщения с помощью ДВП дает хорошие результаты, которые превосходят методы ДКП. Многомасштабный вейвлетный анализ разлагает сигнал в узкие частотные области, что позволяет скрыть сообщение в мелких деталях изображения.

Заключение

В статье были рассмотрены некоторые из основных методов стеганографии изображений, однако существует множество методов сокрытия информации в изображениях. Все основные форматы графических файлов имеют различные методы сокрытия сообщений со своими сильными и слабыми сторонами. Выбор метода с большой надежностью противостоит методу с высокой скоростью обработки. Например, патч-подход имеет очень высокую устойчивость по отношению к большинству видов атак, но он может скрыть лишь очень небольшое количество информации. Поэтому более разумно скрывать информацию в дополнительных преобразованиях, а не в исходных файлах. Преобразование дискретными вейвлетами более надежно, потому что позволяет скрыть сообщение в области частот. Данная область менее подвержена зрению человека.

Авторы предлагают использовать новые методы стеганографии, т. е. алгоритмы вставки скрытого сообщения в изображение с использованием ДВП. Эти методы могут учитывать многосторонние требования пользователя, предъявляемые к качеству изображения со скрытым сообщением большого объема.

Список литературы

1. Currie, D.L. Surmounting the effects of lossy compression on Steganography / D.L. Currie, C.E. Irvine // 19th National Information Systems Security Conference. – Baltimore, 1996.
2. Ahsan, K. Practical Data hiding in TCP/IP / K. Ahsan, D. Kundur // Proceedings of the Workshop on Multimedia Security at ACM Multimedia. – Canada, 2002.
3. Johnson, N.F. Exploring Steganography: Seeing the Unseen / N.F. Johnson, S. Jajodia // Computer Journal. – 1998. – Vol. 31, № 2. – P. 26–34.
4. Katzenbeisser, S. Information Hiding Techniques for Steganography and Digital Watermarking / S. Katzenbeisser, F.A.P. Petitcolas. – Norwood, MA : Artech House, 2000.
5. Niimi, M. High Capacity and Secure Digital Steganography to Palette-Based Images / M. Niimi // IEEE International conference on image processing. – Japan, 2002. – Vol. 2. – P. 917–920.
6. Al-Sadoon, B. On the Development of Steganographic Tools / B. Al-Sadoon, H. Mathkour, Gh. Assassa // Proceedings of the 1st National Information Technology Symposium, NITS. – Riyadh, Saudi Arabia, 2006.
7. Morkel, T. An Overview of Image Steganography / T. Morkel, J.H.P. Eloff, M.S. Olivier. – South Africa : University of Pretoria, 2002.
8. Bender W. Techniques for data hiding / W. Bender, D.Gruhl // IBM Systems Journal. – 1996. – Vol. 35, № 3–4.
9. Krenn, R. Steganography and steganalysis, Internet Publication, March 2004 [Electronic resource] / R. Krenn. – Mode of access : <http://www.krenn.nl/univ/cry/steg/article.pdf>. – Date of access : 24.12.2012.
10. Steganography Software [Electronic resource]. – Mode of access : <http://www.jjtc.com/Steganography/tools.html>. – Date of access : 23.12.2012.
11. Brundick, F. Implementation of Spread Spectrum Image Steganography / F. Brundick, L. Marvel [Electronic resource]. – 2001. – Mode of access : <http://permanent.access.gpo.gov/gpo2618/ARL-TR-2433.pdf>. – Date of access : 24.12.2012.
12. Chang, C.C. Introduction to the Visual Cryptography / C.C. Chang, L.Z. Chuang // Communication of the Chinese Cryptology and Information Security Association (CCISA). – 2004. – Vol. 10, № 2. – P. 1–14.
13. Meenu Kumari, A. JPEG Compression Steganography & Cryptography Using Image-Adaptation Technique / A. Meenu Kumari, P. Khare // Journal of Advances in Information Technology. – 2010. – Vol. 1, № 3.
14. Bushra Kassim Al-Abudi. Colour Image Data Compression Using Multilevel Block Truncation Coding Technique / Bushra Kassim Al-Abudi // Phd Thesis. – Baghdad : College of Science, 2002.
15. Sua'd Kakil, A. Image in Image Hiding System using Iterated Function System (IFS) / A. Sua'd Kakil // Msc Theses. – Iraq : University of Sulaimani, 2009.

16. Khosravi, S. A new steganography method based HIOP (Higher Intensity Of Pixel) algorithm and Strassen's matrix multiplication / S. Khosravi, M.A. Dezfoli, M.H. Yektaie // Journal of Global Research in Computer Science. – 2011. – Vol. 2, № 1.

17. Mathkour, H. A New Image Steganography Technique / H. Mathkour, B. Al-Sadoon, A. Tourir // Wireless Communications, Networking and Mobile Computing (WiCOM) : 4th International Conference on IEEE. – Dalian, China, 2008. – P. 1–4.

18. Babu, S. Authentication of Secret Information in Image Steganography / S. Babu, K.B. Raja // IEEE. – India : University of Heydarabad, 2008. – P. 1–6.

Поступила 27.06.12

*Белорусский государственный университет
информатики и радиоэлектроники,
Минск, ул. П. Бровки, 6
e-mail: amseyyedi@gmail.com
rsadykhov@bsuir.by*

S.A. Seyyedi, R.Kh. Sadykhov

COMPARISON OF DIGITAL IMAGE STEGANOGRAPHY METHODS

Steganography is a method of hiding information in other information of different format (container). There are many steganography techniques with various types of container. In the Internet, digital images are the most popular and frequently used containers. We consider main image steganography techniques and their advantages and disadvantages. We also identify the requirements of a good steganography algorithm and compare various such algorithms.

ПРИКЛАДНЫЕ ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

УДК 004.31

А.А. Лялинский

ОСОБЕННОСТИ ПОСТРОЕНИЯ ПРИКЛАДНЫХ ПРОГРАММ С ВЕБ-ДОСТУПОМ

Рассматриваются вопросы организации веб-интерфейса к модулям, входящим в состав систем автоматизированного проектирования (САПР) в микроэлектронике. Исследуется влияние ряда факторов (доступной памяти компьютера, максимально разрешенного времени выполнения, степени однородности программного обеспечения) на структуру создаваемой системы. Описываются синхронный и асинхронный варианты взаимодействия управляющей и исполнительной частей. Даются ссылки на системы, позволяющие иметь доступ к прикладным программам в САПР микроэлектроники и построенные на принципах, изложенных в статье.

Введение

В микроэлектронике САПР активно используются уже десятки лет. Общепринятым для них является размещение на локальном компьютере пользователя или в локальной сети предприятия. Общение с такой САПР ведется через встроенную систему ввода-вывода данных (также известную как GUI – *Graphical User Interface*), тесно интегрированную с программным обеспечением (далее – специальное ПО), реализующим те задачи, на которые ориентирована данная САПР.

Оснащение систем проектирования веб-доступом, с одной стороны, обеспечивает более широкие возможности по их использованию, существенно расширяя круг лиц, которым эти системы становятся доступны. С другой стороны, с введением веб-доступа возникает ряд проблем юридического и технического характера. Не затрагивая вопросы соблюдения лицензионных соглашений и прав на использование такого ПО, рассмотрим только особенности построения данного специального ПО.

1. Структура связки «веб-сайт – специальное ПО»

Объект «веб-сайт – специальное ПО» (рис. 1) включает следующие части:

- ПО, отвечающее за входной веб-интерфейс и реализованное на предназначенных для этого языках (HTML, PHP, Javascript и т. п.);
- специальное ПО, обеспечивающее собственно вычислительные задачи, для решения которых и создана данная система;
- блок визуализации выходных данных.

Здесь может присутствовать также блок управления, но его роль обычно невелика, так как на нем лежит только задача подготовки пакетного запуска специального ПО.

Интерфейсная часть и блок визуализации размещены на сервере и, в случае использования языков Javascript или Java, частично на клиентской машине. Специальное ПО всегда размещено на сервере.

На ПО веб-интерфейса и блока визуализации лежит ответственность за прием и передачу всех данных между пользователем и специальным ПО. Блок входного интерфейса (или блок управления, если он есть) является управляющим по отношению к блоку специального ПО, т. е. вызовы всегда идут от интерфейсной части к специальной, а не наоборот. Последовательность действий такова:

1. Интерфейсная часть в режиме диалога обеспечивает прием данных от пользователя (по частям или единым пакетом).
2. При поступлении от пользователя сигнала об окончании формирования данных проверяется их реальная полнота и непротиворечивость отдельных частей входного пакета друг другу.

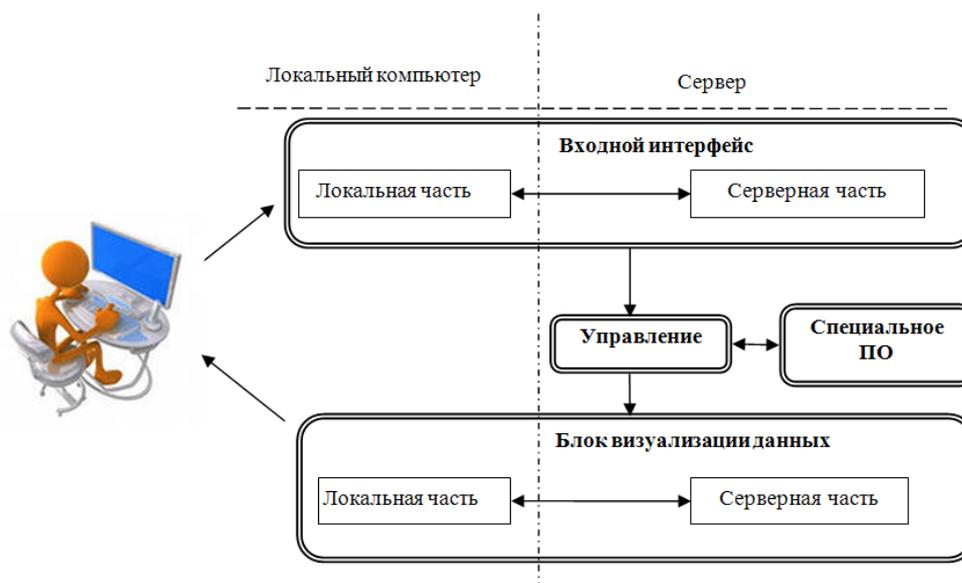


Рис. 1. Структурная схема связки «веб-сайт – специальное ПО»

3. Если получена положительная оценка полноты и правильности входных данных, то формируется внутренний блок данных, поступающий на вход блока специального ПО. Этот блок может быть размещен либо в оперативной памяти сервера, либо, если объем данных велик или непредсказуем, в виде файла на сервере. Если специальное ПО реализовано на скриптовом языке (например, на Матлабе [1, 2]), то при небольшом объеме входных данных можно отказаться от генерации специального файла входных данных, а вместо этого создавать модуль на скриптовом языке, в котором эти данные заданы путем инициализации переменных.

4. От интерфейсной части управление передается на формирование пакета вызова блока специального ПО.

5. Далее возможны два варианта взаимодействия управляющей и исполнительных частей системы. При *синхронном* варианте работа управляющей части приостанавливается до полного завершения работы исполнительных частей. Это более простой вариант, можно сказать, «вариант по умолчанию». В *асинхронном* варианте управляющая часть после запуска части специального ПО начинает выполнение цикла, в котором через определенные промежутки времени происходит опрос флагов состояния блока специального ПО или наличия и/или содержимого определенных файлов на сервере. Выход из цикла возможен либо при получении сигнала об окончании выполнения задачи на стороне процесса специального ПО, либо при превышении заранее определенного лимита времени.

6. При завершении работы исполнительного блока извлекаются результаты его работы (скорее всего, это будет файл на сервере).

7. Если выходные данные блока специального ПО получены, они передаются на вход модуля визуализации данных с последующим его вызовом.

2. Особенности интерфейсной части

Та часть входного интерфейса блока, которая организует прием данных от пользователя, особого интереса в рассматриваемом случае не представляет, так как занимается обработкой информации, полученной из входных форм `<input ... >`, которые заключены в блок (или блоки) `<form> ... </form>`. Следует только обратить внимание на обязательное наличие проверки введенной информации как по линии защиты от хакерских атак, проводимых путем внедрения взламывающего кода в информацию, которая размещается в полях входных форм, так и по линии ее соответствия (как в целом, так и отдельных частей) требованиям, предъявляемым к входной информации данной конкретной задачи.

Рассмотрим организацию управляющего вызова из этого блока модуля специального ПО, который, собственно, и выполняет вычислительную задачу. Различного рода комбинации вызовов вытекают из ограничений, накладываемых на объем оперативной памяти, которая требуется для выполнения специальной части программы, и времени ее выполнения. Кроме того, важную роль играет вопрос об однородности ПО интерфейсной и специальных частей. Далее вкратце рассмотрим влияние этих трех факторов (памяти, времени выполнения, однородности ПО) на структуру системы.

2.1. Ограничение по оперативной памяти

В современных компьютерах требования к объему оперативной памяти, занимаемому выполняющей программой, отошли на второй план. Причина заключается в том, что при ее нехватке происходит сброс части памяти ОЗУ на диск с последующим ее восстановлением при обращении алгоритма к сохраненной на диске части программы (так называемая «виртуализация памяти»). Таким образом, вместо ограничения на память ОЗУ имеем ограничение на объем дискового пространства, отведенного под хранение виртуальных страниц, причем пользователь при необходимости может легко его увеличить. Естественная плата за использование виртуальной памяти – замедление работы программы из-за потерь на обмен страниц памяти между ОЗУ компьютера и диском. Подчеркнем, что главное здесь – отсутствие жесткого ограничения на объем используемой памяти.

К сожалению, при работе скриптового ПО, на котором написано большинство современных сайтов, имеет место реальное фиксированное ограничение на объем памяти, отведенной под динамически размещаемые блоки данных. Так, язык PHP в своем конфигурационном файле `php.ini` имеет параметр `memory_limit`, в котором задается ограничение на допустимый объем динамических данных (8 МБ для версий PHP до 5.2.0, 16 МБ для более старших версий), которые можно разместить при работе исполняемого скрипта. При нехватке памяти программист (создатель скрипта) может попытаться с помощью функции `ini_set(memory_limit, запрашиваемый_размер)` расширить эту границу, что принципиально данную проблему не решает, так как запрашиваемый размер не может быть очень большим.

Отмеченное ограничение имеет значение, если известно, что модуль специального назначения сам по себе невелик и для простоты мог бы быть реализован на сервере непосредственно на одном из скриптовых языков, но во время своей работы он требует больших объемов динамической памяти (например, в задачах обработки образов). В этом случае приходится считаться с объемом доступной динамической памяти.

Итак, при реализации специальной части на скриптовых языках следует помнить о существовании жесткого ограничения на доступную динамическую память.

2.2. Ограничение времени выполнения

При запуске задачи на локальном компьютере (или на любом компьютере, доступном в локальной сети) ограничения по времени выполнения задачи практически нет. Для больших задач реально оно определяется терпением пользователя или истощением дискового пространства выходными данными, что опять-таки не является жестким ограничением.

К сожалению, ограничение по времени имеет место при запуске задачи на веб-сервере. Причина заключается в том, что связь между клиентом, дающим команду со своего локального компьютера на запуск задачи, и веб-сервером, на котором размещена задача специального вида, осуществляется через канал, представляющий собой совокупность программных средств и физических линий связи. Так как пропускная способность канала не безгранична, существует конкретное ограничение по максимальному времени ожидания отклика от сервера (*timeout limit*). Это ограничение повышает пропускную способность канала, так как позволяет выводить из оборота «повисшие» запросы к серверу. Реально речь идет об отрезке времени в 30 с. Можно попытаться расширить этот временной диапазон изменением настроек веб-сервера, что принципиально проблему не решает.

Итак, второе ограничение на специальное ПО – малое время выполнения. Позже увидим, что это ограничение можно преодолеть.

2.3. Однородность интерфейсной и специальной частей

Интерфейсная часть, отвечающая за непосредственный контакт с пользователем, обычно представляет собой набор модулей, написанных на статических языках описания (HTML, XML), и поведенческих модулей, написанных на каких-либо скриптовых языках (PHP, JavaScript, ASP, Perl и т. п.). Скриптовые языки обычно исполняются в режиме интерпретации. Существует возможность написать интерфейсную часть и на компилируемых языках (С, С++, С#), но в данном случае это не представляет интереса из-за большей трудности написания (по сравнению со скриптовыми языками), малой распространенности и из-за того, что канал связи всегда работает медленнее компьютера и не позволяет воспользоваться ускорением, полученным при использовании компилируемых языков.

Специальное ПО обычно уже существует в каком-либо виде к тому моменту, когда возникнет потребность в подключении к нему веб-доступа. Вариантов программной реализации специального ПО великое множество, но в данном случае существенно только то, написано оно в виде скриптов или в виде компилируемых модулей. Если специальное ПО написано в виде скриптов, то проблем с разнородностью скриптов не будет, данные между управляющим и исполнительным скриптами могут передаваться через механизмы глобальных переменных `$_POST`, `$_GET`, `$_SESSION`, управление между скриптами осуществляется посредством http-вызовов. Если же специальное ПО написано на компилируемом языке (или на скриптовом, но выполняемом в режиме компиляции MATLAB), то такой гибкости уже нет. Данные в специальную часть и обратно передаются посредством файлов, управление обычно одностороннее – только в виде вызовов от интерфейсной части к специальной. Поэтому для небольших задач, написанных в исходном виде на компилируемом языке, имеет смысл рассмотреть вопрос о целесообразности их перевода на скриптовый язык.

3. Взаимодействие частей системы

Введем обобщенное определение управляющей и исполнительной частей системы. *Управляющая* часть включает в себя ПО, размещенное на сервере и, возможно, на локальном компьютере клиента и отвечающее как за организацию диалогов для приемки входных параметров, так и за обеспечение визуализации выходных данных. *Исполнительная* часть размещена только на сервере и отвечает за вычислительные аспекты решения задачи.

Как уже упоминалось ранее, возможны два варианта взаимодействия управляющей и исполнительной частей системы: *синхронный* и *асинхронный* (рис. 2).

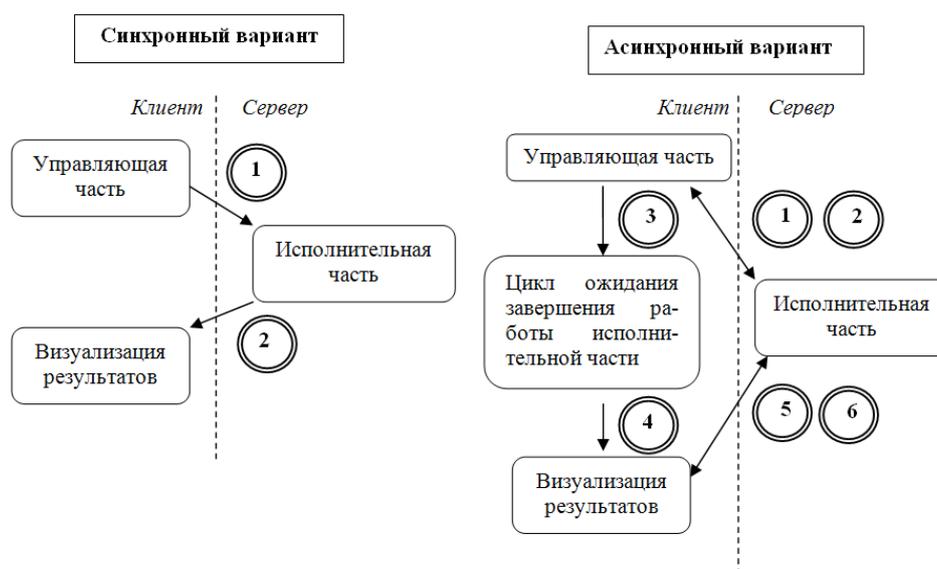


Рис. 2. Варианты взаимодействия управляющей и исполнительной частей

Синхронный вариант предполагает обычную последовательность работ: управляющая часть выдает задание исполнительской (шаг 1) и приостанавливает свою работу в ожидании сигнала завершения вычислительной задачи. По завершении работы исполнительской части происходит визуализация полученных результатов (шаг 2). Какие-либо средства синхронизации задач не применяются.

При выборе асинхронного варианта принципиальной является одновременная работа обеих частей системы и использование специальных методов для синхронизации их работы. Управляющая часть посылает запрос на исполнение (шаг 1) и сразу же возвращается к себе (шаг 2), запуская цикл ожидания завершения работы исполнительской части (шаг 3). При получении сигнала о завершении работы вызывается блок визуализации (шаг 4), который запрашивает (шаг 5) и получает (шаг 6) данные. Преимущество асинхронного подхода заключается в том, что он позволяет преодолеть ограничение на время работы исполнительской части программы (так как управляющая часть не держит канал занятым только для того, чтобы узнать об окончании работы исполнительской части – вместо этого проводится периодический опрос флагов состояния).

Использование синхронного варианта работы возможно (и предпочтительно), если известно, что время работы исполнительской части не превышает установленный максимальный таймаут. В этом случае можно ожидать, что соединение клиента с сервером не будет разорвано из-за превышения таймаута и система в целом обеспечит устойчивую работу. В противном случае необходимо использовать асинхронные методы.

В настоящее время на веб-серверах наиболее распространена организация асинхронной работы на основе технологии AJAX [3, 4]. Рассмотрим пример ее работы.

3.1. Схема работы с использованием AJAX-технологии

AJAX-технология предполагает написание на языке Javascript модуля, в котором на стороне клиента выполняются следующие действия:

- создание XMLHttpRequest-объекта;
- подготовка запроса;
- указание функции, отслеживающей состояние запроса;
- отправка запроса.

Создание XMLHttpRequest-объекта. Объект XMLHttpRequest дает возможность браузеру выполнять запросы к серверу без перезагрузки страницы. Кроссбраузерный текст создания XMLHttpRequest-объекта на языке javascript имеет вид

```
// code for IE7+, Firefox, Chrome, Opera, Safari
var xmlhttp;
if (window.XMLHttpRequest)
    xmlhttp = new XMLHttpRequest();
// code for IE6, IE5
else
    xmlhttp = new ActiveXObject("Microsoft.XMLHTTP");
```

Подготовка запроса. Для подготовки запроса используется метод open объекта XMLHttpRequest. Имеется несколько вариантов списков параметров данного метода, наиболее простой для асинхронного вызова имеет вид

```
xmlhttp.open (тип запроса, url);
```

где тип запроса – 'GET' или 'POST'. В GET-запросе параметры передаются в строке вызова исполняемого скрипта, в POST-запросе – в теле, посылаемом через метод *send* (см. далее). Объем

параметров при посылке с использованием POST неограничен. Для POST необходим заголовок Content-Type, содержащий кодировку

```
xmlHttpRequest.setRequestHeader ('Content-Type', 'application/x-www-form-urlencoded');
```

url – адрес вызываемого скрипта.

Указание функции, отслеживающей состояние запроса. При асинхронном вызове в методе onreadystatechange необходимо указать функцию, которая отслеживает состояние объекта и вызывается каждый раз, когда XMLHttpRequest-объект изменяет состояние. Реально из пяти возможных состояний (0 – Uninitialized, 1 – Loading, 2 – Loaded, 3 – Interactive, 4 – Complete) гарантированно можно полагаться только на получение последнего (с кодом 4). Получение остальных зависит от браузера, использовать их не рекомендуется. Кроме проверки состояния XMLHttpRequest-объекта рекомендуется также проверка состояния запроса, возможные ответы: 200 – ОК, 404 – Page not found. В результате минимально возможное описание функции имеет вид

```
xmlHttpRequest.onreadystatechange = function() {
    if (xmlHttpRequest.readyState == 4) {
        if (xmlHttpRequest.status == 200) {
            alert(xmlHttpRequest.responseText);
        }
    }
}
```

Отправка запроса. Для отправки запроса используется метод send объекта XMLHttpRequest. При GET-запросе в качестве параметра указывается null, при POST-запросе – список параметров, например

```
xmlHttpRequest.send ("s=5&b=3");
```

После выполнения данной команды происходит непосредственный запуск асинхронно выполняемого скрипта.

3.2. Проблема индикации работы выполняемого блока

При выполнении достаточно длительных задач желательно показывать пользователю степень их выполнения. Учитывая итерационный характер решаемых в САПР микроэлектроники задач, выбрать для показа нужный индикатор нетрудно. К сожалению, в связке «веб-сайт – специальное ПО» возникает проблема технического характера, препятствующая такому показу.

При выполнении задачи в *синхронном* режиме управление полностью передается блоку специального ПО, которое не имеет возможности выводить какие-либо данные на экран компьютера пользователя, так как выполняется в пакетном режиме и вся работа с входными-выходными данными ведется через файлы.

В *асинхронном* режиме одновременно работают и интерфейсная часть, и блок специального ПО. Последний мог бы выводить информацию о своем состоянии в файл, но проблема в том, что в AJAX-технологии управляющая часть написана на javascript и, следовательно, размещена на компьютере клиента. По определению языка javascript в нем в целях безопасности не предусмотрены средства работы с файлами на сервере, поэтому даже при наличии файла состояния задачи нет возможности вывода данных из него на экран.

Таким образом, можно показывать только состояния запуска и окончания выполнения задачи.

4. Визуализация данных

Известно, что основные скриптовые языки, на которых создается большинство веб-сайтов, в своем составе не содержат графических средств, позволяющих проводить полноценную визуализацию данных (строить графики, диаграммы и т. п.). К счастью, в настоящее время существуют графические пакеты различного уровня, позволяющие выполнить такую работу. На сайте документации по PHP5 [5] (наиболее широко используемой версии PHP в настоящее время) доступна информация (есть полное описание, список функций и руководство по установке) по трем графическим пакетам: Cairo [6], GD [7] и ImageMagick [8]. Эти графические пакеты позволяют проводить построение основных геометрических фигур, проводить их перемещение, раскраску, выполнять операции с альфа-каналом, читать и сохранять полученный образ в наиболее известных графических типах данных.

Работу с данными на более высоком уровне предоставляет графический пакет JpGraph [9], который дает возможность построения огромного количества типов графиков и диаграмм. Для бесплатной версии состав доступных функций ограничен, так же, как и возможность ее использования в коммерческих версиях.

Можно также упомянуть бесплатный пакет pChart [10], являющийся надстройкой над GD-библиотекой и предоставляющий средства для отображения различного рода статистической информации.

5. Примеры реализации прикладных программ с веб-доступом

На основе проведенного по данной тематике анализа в ИППМ РАН были созданы следующие системы [11–13]:

1. Система расчета *spice*-макромодели интегральной спиральной индуктивности, позволяющая рассчитывать сосредоточенные параметры индуктивности (собственно индуктивность, межвитковую емкость, сопротивление дорожек), формируемой на кристалле интегральной схемы по геометрическим размерам ее топологии и заданной технологии.

Исполнительная часть реализована в виде *exe*-модуля, исходный код которого написан на языке C.

2. Система определения оптимального набора тестов логической функции, позволяющая создавать оптимальный (с точки зрения последующего *spice*-моделирования) набор входных тестов для произвольной логической функции. Создаваемый набор тестов имеет минимально возможное количество входных воздействий.

Исполнительная часть написана полностью на PHP.

3. Система автоматической генерации *Verilog*-моделей, позволяющая создавать высокоуровневые модели (на языках *Verilog* или *Verilog-AMS*) цифровых комбинационных схем на основе заданного пользователем произвольного логического выражения. Кроме того, при указании технологической *spice*-библиотеки ячеек возможно построение моделей с временными параметрами. Это дает возможность упростить и автоматизировать процесс разработки высокоуровневых моделей цифровых ячеек, что, в свою очередь, ускоряет разработку крупных проектов в целом.

Исполнительная часть написана на PHP, кроме того, возможен вызов *exe*-модуля *spice*-симулятора.

4. Система моделирования инерционной навигационной системы, которая позволяет навигатору в течение некоторого времени после потери контакта со спутником продолжать ориентирование на местности исходя из показаний встроенных в него датчиков движения, а также оценивать точность работы инерционной навигационной системы и степень доверия к выдаваемым данным в зависимости от используемых сенсоров движения и параметров настройки ПО.

Исполнительная часть написана на языке Matlab.

В первых трех системах исполнительная часть вызывается синхронно, в последней – асинхронно.

Заключение

Оснащение систем автоматизированного проектирования веб-доступом значительно расширяет возможности по их применению, предоставляя к ним доступ из любой точки, где есть Интернет. Относительно вопроса о том, можно ли добавить веб-доступ к реально существующим САПР-системам, следует признать, что в настоящее время наиболее реально построение САПР с веб-доступом на основе специально написанных (или переработанных) блоков, учитывающих специфику требований, рассмотренных в данной статье.

Список литературы

1. MATLAB – пакет прикладных программ для решения задач технических вычислений и одноименный язык программирования [Электронный ресурс]. – Режим доступа : <http://ru.wikipedia.org/wiki/MATLAB>. – Дата доступа : 10.12.2012.
2. Дьяконов, В.П. MATLAB R2006/2007/2008 + Simulink 5/6/7. Основы применения / В.П. Дьяконов. – Изд-е 2-е, перераб. и доп. – М. : СОЛОН-Пресс, 2008. – С. 800. – (Библиотека профессионала).
3. AJAX (*Asynchronous Javascript and XML*) [Электронный ресурс]. – Режим доступа : <http://ru.wikipedia.org/wiki/AJAX>. – Дата доступа : 10.12.2012.
4. Крейн, Д. AJAX в действии: технология – *Asynchronous JavaScript and XML = Ajax in Action* / Д. Крейн, Э. Паскарелло, Д. Джеймс. – М. : Вильямс, 2006. – 640 с.
5. Руководство по PHP [Электронный ресурс]. – Режим доступа : <http://php.net/manual/ru/index.php>. – Дата доступа : 10.12.2012.
6. Cairo: 2D graphics library [Электронный ресурс]. – Режим доступа : <http://cairographics.org/>. – Дата доступа : 10.12.2012.
7. GD Graphics Library [Электронный ресурс]. – Режим доступа : http://ru.wikipedia.org/wiki/GD_Graphics_Library. – Дата доступа : 10.12.2012.
8. ImageMagick [Электронный ресурс]. – Режим доступа : <http://www.imagemagick.org/script/index.php>. – Дата доступа : 10.12.2012.
9. JpGraph: an Object-Oriented Graph creating library for PHP [Электронный ресурс]. – Режим доступа : <http://jgraph.net/>. – Дата доступа : 10.12.2012.
10. pChart [Электронный ресурс]. – Режим доступа : <http://www.pchart.net/>. – Дата доступа : 10.12.2012.
11. Лялинский, А.А. Формирование высокоуровневых моделей цифровых ячеек с использованием веб-доступа // Проблемы разработки перспективных микро- и наноэлектронных систем – 2012 : сб. тр. / под общ. ред. акад. РАН А.Л. Стемпковского. – М. : ИППМ РАН, 2012. – С. 101–106.
12. Лялинский, А.А. Автоматизированное формирование тестов при характеристике цифровых ячеек с использованием веб-доступа / А.А. Лялинский // Проблемы разработки перспективных микро- и наноэлектронных систем – 2012 : сб. тр. / под общ. ред. акад. РАН А.Л. Стемпковского. – М. : ИППМ РАН, 2012. – С. 95–100.
13. Институт проблем проектирования в микроэлектронике РАН. Раздел Online-проектирования [Электронный ресурс]. – Режим доступа : <http://www.ippm.ru/index.php?page=webcad>. – Дата доступа : 10.12.2012.

Поступила 18.12.2012

*Институт проблем проектирования
в микроэлектронике РАН,
Москва, Зеленоград,
ул. Советская, 3
e-mail: zelyal@inbox.ru*

A.A. Lyalinsky

WEBSITE EXECUTION OF CAD MODULES

Web-based interface of modules that are part of the computer-aided design system in microelectronics is considered. The influence of several factors (available computer memory, maximum allowed run time, degree of homogeneity of the software) on the structure of the created system is investigated. Synchronous and asynchronous variants of interaction between control and executive parts are described. References on the systems that allow an access to applications in CAD microelectronics and are based on the principles discussed in this article are given.

УДК 796.6+519.2

И.Ю. Михута

АЛГОРИТМ ОЦЕНКИ И ПРОГНОЗИРОВАНИЯ ПСИХОФИЗИЧЕСКОЙ ГОТОВНОСТИ И ПРИГОДНОСТИ УЧАЩЕЙСЯ МОЛОДЕЖИ К ВОЕННО-ПРОФЕССИОНАЛЬНОЙ ДЕЯТЕЛЬНОСТИ

Рассматривается задача разработки алгоритма оценки и прогнозирования психофизической готовности и пригодности учащейся молодежи к военно-профессиональной деятельности. Предлагаемый статистический алгоритм позволяет обеспечить надежность и информативность методики диагностики, обоснованность качественных и количественных критериев оценки, простоту и доступность вычисления, а также применимость матрицы психофизических критериев готовности и пригодности субъектов к конкретным профилям (экстремальному, операторско-инженерному и общевойсковому) военно-профессиональной деятельности.

Введение

Основным направлением модернизации Вооруженных Сил Республики Беларусь является оптимизация использования человеческих ресурсов, что требует оценки готовности и пригодности военного специалиста к выполнению той или иной военно-профессиональной деятельности. При проведении оценки и прогнозирования профессиональной психофизической готовности и пригодности кандидатов к военно-профессиональной деятельности необходимо опираться на системно-структурный анализ личности, основные принципы обследования, методологию диагностики, организационные формы проведения отбора, а также на связь результатов определения готовности и пригодности с другими формами обеспечения успешности обучения и эффективности военно-профессиональной деятельности [1–4].

По мнению В.А. Бодрова, В.Я. Апчела, особое значение для оценки и прогнозирования профессиональной пригодности имеют такие принципы профотбора, как целенаправленность изучения личности, комплексность оценки профессиональной пригодности, личностно-деятельностный принцип, объективность и обоснованность оценки, этапность, группировка специальностей (дифференцированное прогнозирование), практичность. С учетом этих принципов должны разрабатываться критерии оценки и прогнозирования профессиональной готовности и пригодности молодежи к военно-профессиональной деятельности, развиваться и совершенствоваться система профессионального отбора и диагностики будущих специалистов военной сферы [5, 6].

Необходимость проведения мероприятий по выяснению профессиональной готовности и пригодности обусловлена индивидуальными различиями в профессионально важных психофизических характеристиках будущих специалистов военного профиля, а также наличием зависимости между этими характеристиками.

Процедура оценки и прогнозирования профессиональной готовности и пригодности, в частности ее методический и критериальный аппарат, в значительной степени зависит от содержания задач профотбора. При отборе по конкретным видам военно-профессиональной деятельности в качестве внешних критериев пригодности выступают показатели эффективности, надежности, безопасности трудового процесса, стремление к профессиональному совершенствованию и т. п. При постановке этой задачи изучению и оценке подлежат психофизические способности и другие качества личности, обеспечивающие необходимый уровень достижений в профессиональной, социальной и личностной сферах.

Особую значимость военно-профессиональный отбор имеет на начальных этапах профессионального самоопределения учащихся и курсантов в военных учебных заведениях, когда на основе дифференцированной оценки и прогнозирования готовности и пригодности к ряду специальностей военно-профессионального профиля можно определить ту специальность, которая по своим психофизическим требованиям наиболее точно соответствует психофизическим особенностям выпускника. Процедура оценки и прогнозирования профессиональной готовности и пригодности в этом случае основана, с одной стороны, на необходимости соотнесения

индивидуально-психофизических особенностей личности с набором профессиональных требований соответствующих специальностей, а с другой – на возможности использования дополнительной психофизической информации (матрицы психофизической готовности и пригодности), полученной при обследовании на ранних стадиях обучения, которая необходима для повышения надежности прогноза будущей военно-профессиональной деятельности [2].

Так, в прогнозировании готовности и пригодности будущих специалистов в военно-профессиональной деятельности предлагается выделить пять стадий и этапов организации непосредственной оценки.

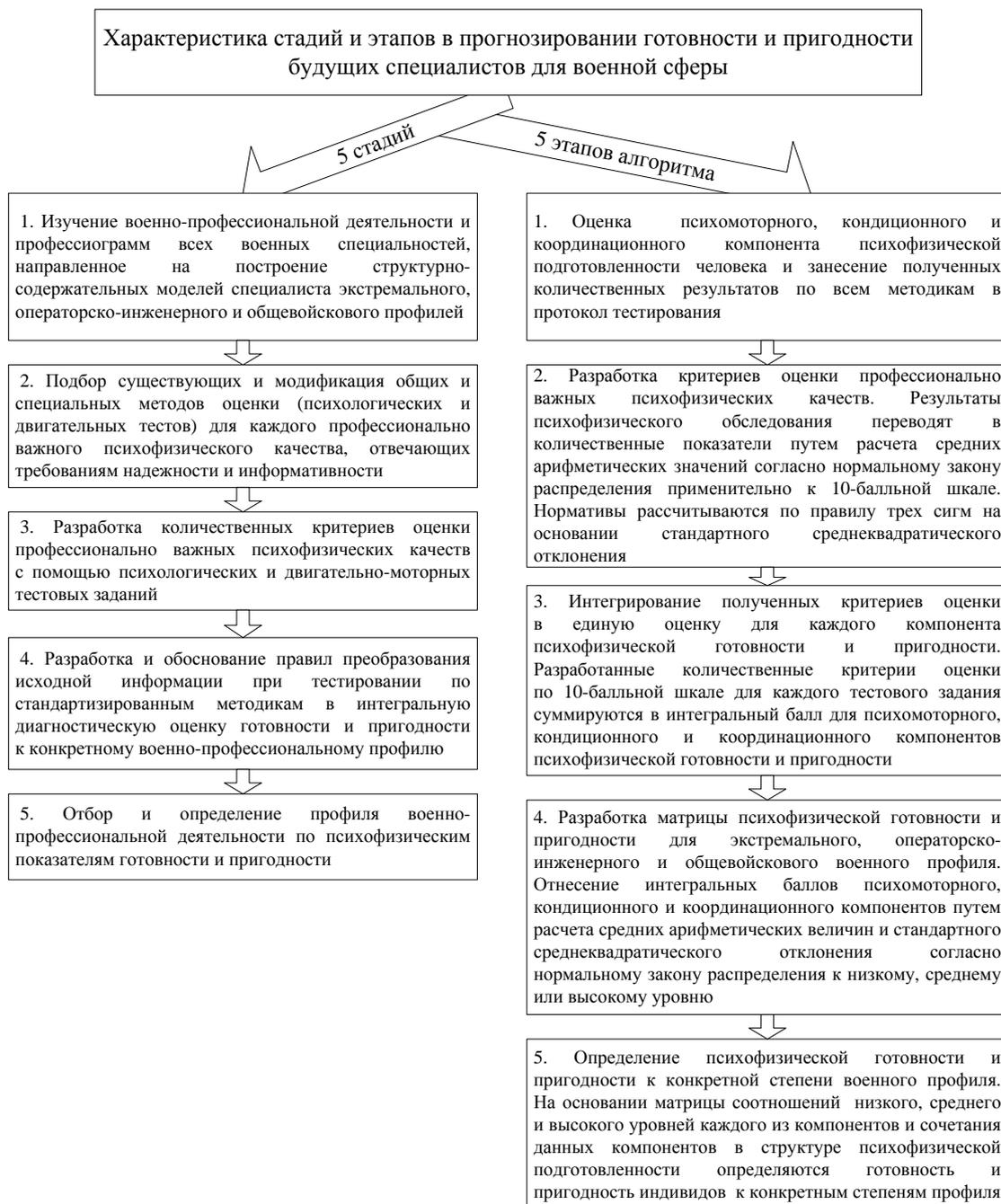


Рис. 1. Основные стадии и этапы прогнозирования психофизической готовности и пригодности учащейся молодежи к военно-профессиональной деятельности

В настоящее время известны различные аналитические и статистические приемы получения и обработки результатов психофизического обследования пригодности кандидатов к тем или иным видам профессиональной и трудовой деятельности [1, 7]. Однако в современной пси-

холого-педагогической литературе отсутствуют работы, в которых рассматривается алгоритмизация оценочных критериев готовности и пригодности молодежи к будущей военно-профессиональной деятельности, особенно на начальном этапе профессионального самоопределения. В настоящей статье предлагается статистический алгоритм оценки психофизической готовности и пригодности человека для прогнозируемых специальностей и профилей будущей военно-профессиональной деятельности.

1. Постановка задачи

Пусть информация о психофизических способностях обследуемых содержится в n -мерном векторе $v=(v_1, v_2, \dots, v_n)$. Каждое число $v_i, i = 1, 2, \dots, n$, получено при помощи некоторой методики тестирования. В этих числах могут быть закодированы и другие качественные характеристики индивида. В дальнейшем компоненты v будут называться признаками. Выбор признаков обычно производится с учетом психофизических требований к военно-профессиональной готовности и пригодности. Предлагаемый алгоритм позволяет отбросить те из используемых признаков, которые оказываются неинформативными.

Предполагается, что существует три группы лиц, пригодных к военно-профессиональной деятельности: экстремального профиля (ЭП); операторско-инженерного профиля (ОИП); общевойскового профиля (ОВП). Им соответствуют три класса $\Omega_{ЭП}, \Omega_{ОИП}$ и $\Omega_{ОВП}$ n -мерных векторов. Эти классы могут пересекаться, но статистически имеют различия.

Задача определения профессиональной готовности и пригодности заключается в отнесении с определенной (малой) вероятностью ошибки вектора (v_1, v_2, \dots, v_n) к одному из трех классов: $\Omega_{ЭП}, \Omega_{ОИП}$ и $\Omega_{ОВП}$.

Успех классификации по многим признакам в задачах диагностики и отбора зависит от информативности этих признаков и способа интеграции информации. Способ интеграции должен быть [7]:

- простым в вычислительном отношении и доступным при использовании;
- малочувствительным к отсутствию какого-либо признака;
- достаточно инвариантным к сдвигу распределений признаков.

Поскольку психофизические признаки v_1, v_2, \dots, v_n являются статистически зависимыми, последовательный анализ служит эффективным подходом для классификации трех профилей военно-профессиональной деятельности.

2. Алгоритм оценки и прогнозирования психофизической готовности и пригодности

Алгоритм состоит из этапа обучения, на котором накапливается информация о признаках на основании уже имеющегося опыта и оценивается информативность выбранных признаков, и этапа классификации, на котором выносится решение о готовности и пригодности субъекта к определенной военно-профессиональной деятельности [4, 8].

2.1. Этап обучения

Предполагается, что на основании проведенного обследования можно выделить группы субъектов выборки $A_{ЭП}, A_{ОИП}$ и $A_{ОВП}$, относящихся к классам $\Omega_{ЭП}, \Omega_{ОИП}$ и $\Omega_{ОВП}$. Вопросы, связанные с образованием выборок по классам $\Omega_{ЭП}, \Omega_{ОИП}$ и $\Omega_{ОВП}$, будут рассмотрены ниже.

Процесс обучения состоит в получении оценки дискретных одномерных распределений вероятностей признаков v_1, v_2, \dots, v_n субъектов по следующим профилям:

- экстремальному

$$f^1_{ЭП}(v_1), f^2_{ЭП}(v_2), \dots, f^n_{ЭП}(v_n); \quad (1)$$

- операторско-инженерному

$$f^1_{ОИП}(v_1), f^2_{ОИП}(v_2), \dots, f^n_{ОИП}(v_n); \quad (2)$$

- общевойсковому

$$f^1_{ОВП}(v_1), f^2_{ОВП}(v_2), \dots, f^n_{ОВП}(v_n). \quad (3)$$

Предполагается, что зависимость v_1, v_2, \dots, v_n пренебрежимо мала. Если, однако, это предположение не выполнено, для увеличения эффективности процедуры в рассмотрение вводятся сложные признаки – синдромы, определение которых можно получить на основании опыта или же используя соответствующий математический аппарат. Построение одномерных распределений существенно облегчает процесс обучения, а в случае слабой зависимости потери информации при этом невелики.

Если выборки $A_{ЭП}$, $A_{ОИП}$ и $A_{ОВП}$ по количеству обследуемых субъектов многочисленны, можно получить достаточно хорошую оценку требуемых вероятностей:

$$\{f^{j_{ЭП}}(v_j)\}, \{f^{j_{ОИП}}(v_j)\}, \{f^{j_{ОВП}}(v_j)\}, \quad j = 1, 2, \dots, n. \quad (4)$$

Полученные в результате обследования показатели могут иметь различную ценность для целей прогнозирования военно-профессиональной готовности и пригодности. Поэтому следующим шагом на этапе обучения является оценка информативности признаков.

Признак будет тем информативнее, чем больше различие между его распределениями у представителей классов $\Omega_{ЭП}$, $\Omega_{ОИП}$ и $\Omega_{ОВП}$. Оценка информативности признака v_j может выражаться оценкой вероятности того, что распределения $f^{j_{ЭП}}(v_j)$, $f^{j_{ОИП}}(v_j)$ и $f^{j_{ОВП}}(v_j)$ различны. Необходимость такой оценки исследуется для данной классификации при помощи статистического критерия хи-квадрат, который позволяет сравнивать распределения частот вне зависимости от того, распределены они нормально или нет [8, 9]. Необходимо отметить, что признаки, информативные для одного вида деятельности, могут оказаться неинформативными для других видов деятельности.

В расчетах используются таблицы Л. Большева и Н. Смирнова [6]. Оценка информативности может быть также получена и при помощи расстояния Кульбака [8]. В обозначениях, принятых в [8], в несколько измененной форме расстояние Кульбака I_j признака v_j представлено в следующем виде:

$$I_j = |I_j^{ЭП}| + |I_j^{ОИП}| + |I_j^{ОВП}|, \quad j = 1, 2, \dots, n; \quad (5)$$

$$I_j^{ЭП} = \sum_{i=1}^{S_j} \left(\frac{ЭП_i^j}{N_{ЭП}^j} + \frac{ЭП_i^j \cdot N_{ОИП}^j \cdot N_{ОВП}^j}{ОИП_i^j \cdot ОВП_i^j \cdot N_{ЭП}^j} \right); \quad (6)$$

$$I_j^{ОИП} = \sum_{i=1}^{S_j} \left(\frac{ОИП_i^j}{N_{ОИП}^j} + \frac{ОИП_i^j \cdot N_{ЭП}^j \cdot N_{ОВП}^j}{ЭП_i^j \cdot ОВП_i^j \cdot N_{ОИП}^j} \right); \quad (7)$$

$$I_j^{ОВП} = \sum_{i=1}^{S_j} \left(\frac{ОВП_i^j}{N_{ОВП}^j} + \frac{ОВП_i^j \cdot N_{ЭП}^j \cdot N_{ОИП}^j}{ЭП_i^j \cdot ОИП_i^j \cdot N_{ОВП}^j} \right), \quad (8)$$

где S_j – количество интервалов значений признаков v_j ; $ЭП_i^j$ – количество субъектов ЭП в i -м интервале; $ОВП_i^j$ – количество субъектов ОИП в i -м интервале; $ОВП_i^j$ – количество субъектов ОВП в i -м интервале; $N_{ЭП}^j$, $N_{ОИП}^j$ и $N_{ОВП}^j$ – общее количество субъектов; $I_j^{ЭП}$ – расстояние Кульбака признака v_j для $\Omega_{ЭП}$; $I_j^{ОИП}$ – расстояние Кульбака признака v_j для $\Omega_{ОИП}$; $I_j^{ОВП}$ – расстояние Кульбака признака v_j для $\Omega_{ОВП}$.

Такая мера имеет ряд преимуществ, особенно при аналитических расчетах. Для практики представляет интерес возможность измерения значимости признаков v_j , $j = 1, 2, \dots, n$, отдельно для вынесения решения о принадлежности v к $\Omega_{ЭП}$, $\Omega_{ОИП}$ и $\Omega_{ОВП}$ (соответственно слагаемым $I_j^{ЭП}$, $I_j^{ОИП}$ и $I_j^{ОВП}$).

Этап алгоритма обучения можно считать законченным, когда оценки распределений $f_j^{\text{ЭП}}(vj)$, $f_j^{\text{ОИП}}(vj)$ и $f_j^{\text{ОВП}}(vj)$, $j = 1, 2, \dots, n$, построены, а признаки упорядочены по их информативности.

2.2. Этап классификации (решающее правило)

Существует шесть типов ошибок при классификации субъекта на три класса $\Omega_{\text{ЭП}}$, $\Omega_{\text{ОИП}}$ и $\Omega_{\text{ОВП}}$:

– субъект из класса $\Omega_{\text{ЭП}}$ ошибочно отнесен в класс $\Omega_{\text{ОИП}}$ или $\Omega_{\text{ОВП}}$ (вероятность ошибки $\alpha_{\text{ОИП}}$ или $\alpha_{\text{ОВП}}$);

– субъект из класса $\Omega_{\text{ОИП}}$ ошибочно отнесен в класс $\Omega_{\text{ЭП}}$ или $\Omega_{\text{ОВП}}$ (вероятность ошибки $\beta_{\text{ЭП}}$ или $\beta_{\text{ОВП}}$);

– субъект из класса $\Omega_{\text{ОВП}}$ ошибочно отнесен в класс $\Omega_{\text{ЭП}}$ или $\Omega_{\text{ОИП}}$ (вероятность ошибки $\gamma_{\text{ЭП}}$ или $\gamma_{\text{ОИП}}$).

Введем следующие обозначения:

$\alpha = \alpha_{\text{ОИП}} + \alpha_{\text{ОВП}}$ – вероятность ошибки классификации субъектов класса $\Omega_{\text{ЭП}}$;

$\beta = \beta_{\text{ЭП}} + \beta_{\text{ОВП}}$ – вероятность ошибки классификации субъектов класса $\Omega_{\text{ОИП}}$;

$\gamma = \gamma_{\text{ЭП}} + \gamma_{\text{ОИП}}$ – вероятность ошибки классификации субъектов класса $\Omega_{\text{ОВП}}$.

Максимально допустимые значения α , β и γ вероятностей ошибок определяются до проведения классификации. При выборе этих вероятностей должна быть учтена важность той или другой ошибки классификации, а также реальная ситуация, возникшая при решении рассматриваемой задачи [9].

Пусть при обследовании субъекта S были получены признаки $v_1^0, v_2^0, \dots, v_n^0$ (они перенумерованы и уже приведены здесь в порядке их убывающей информативности). Рассмотрим отношения вероятностей, соответствующие первому признаку:

$$\frac{f_{\text{ОИП}}^1(v_1^0)}{f_{\text{ЭП}}^1(v_1^0)}, \frac{f_{\text{ОВП}}^1(v_1^0)}{f_{\text{ЭП}}^1(v_1^0)}, \frac{f_{\text{ЭП}}^1(v_1^0)}{f_{\text{ОИП}}^1(v_1^0)}, \frac{f_{\text{ОВП}}^1(v_1^0)}{f_{\text{ОИП}}^1(v_1^0)}, \frac{f_{\text{ЭП}}^1(v_1^0)}{f_{\text{ОВП}}^1(v_1^0)}, \frac{f_{\text{ОИП}}^1(v_1^0)}{f_{\text{ОВП}}^1(v_1^0)}. \quad (9)$$

Если отношения из первой пары вероятностей удовлетворяют неравенствам

$$\frac{f_{\text{ЭП}}^1(v_1^0)}{f_{\text{ОИП}}^1(v_1^0)} < \frac{\beta_{\text{ЭП}}}{1 - \alpha_{\text{ОИП}}}, \frac{f_{\text{ЭП}}^1(v_1^0)}{f_{\text{ОВП}}^1(v_1^0)} < \frac{\gamma_{\text{ЭП}}}{1 - \alpha_{\text{ОВП}}}, \quad (10)$$

то полученное значение признака v_1^0 настолько вероятнее для класса $\Omega_{\text{ЭП}}$, что можно с выбранным уровнем надежности (α , β и γ) утверждать, что данный субъект относится к классу $\Omega_{\text{ЭП}}$. Если же для отношений из первой пары вероятности верны неравенства

$$\frac{f_{\text{ОИП}}^1(v_1^0)}{f_{\text{ЭП}}^1(v_1^0)} > \frac{\alpha_{\text{ОИП}}}{1 - \beta_{\text{ЭП}}}, \frac{f_{\text{ОВП}}^1(v_1^0)}{f_{\text{ЭП}}^1(v_1^0)} > \frac{\alpha_{\text{ОВП}}}{1 - \gamma_{\text{ЭП}}}, \quad (11)$$

то с тем же уровнем надежности принимается решение о непригодности к рассматриваемому экстремальному профилю военно-профессиональной деятельности. Далее целесообразно проверить лишь оставшиеся два класса $\Omega_{\text{ОИП}}$ и $\Omega_{\text{ОВП}}$.

Если выполняется неравенство

$$\frac{f_{\text{ОВП}}^1(v_1^0)}{f_{\text{ОИП}}^1(v_1^0)} < \frac{\beta_{\text{ОВП}}}{1 - \gamma_{\text{ОИП}}}, \quad (12)$$

то субъект относится к классу $\Omega_{\text{ОИП}}$, если выполняется

$$\frac{f_{\text{ОВП}}^1(v_1^0)}{f_{\text{ОИП}}^1(v_1^0)} > \frac{1 - \beta_{\text{ОВП}}}{\gamma_{\text{ОИП}}}, \quad (13)$$

то субъект относится к классу $\Omega_{\text{ОВП}}$. В оставшихся случаях

$$\frac{\beta_{ОВП}}{1-\gamma_{ОИП}} \leq \frac{f_{ОВП}^1(v_1^0)}{f_{ОИП}^1(v_1^0)} \leq \frac{1-\beta_{ОВП}}{\gamma_{ОИП}}, \quad (14)$$

т. е. при имеющейся в признаке недостаточной информации для обеспечения заданной точности следует аналогичным образом рассчитывать отношения из формулы (10).

Если выполняются неравенства

$$\frac{\alpha_{ОИП}}{1-\beta_{ЭП}} \leq \frac{f_{ОИП}^1(v_1^0)}{f_{ЭП}^1(v_1^0)} \leq \frac{1-\alpha_{ОИП}}{\beta_{ЭП}}, \quad \frac{\alpha_{ОВП}}{1-\gamma_{ЭП}} \leq \frac{f_{ОВП}^1(v_1^0)}{f_{ЭП}^1(v_1^0)} \leq \frac{1-\alpha_{ОВП}}{\gamma_{ЭП}}, \quad (15)$$

то информация, заключенная в признаке, недостаточна для отнесения к одному из классов $\Omega_{ЭП}$, $\Omega_{ОИП}$ и $\Omega_{ОВП}$ или исключения одного из этих классов из рассмотрения. В таком случае требуется учитывать значения второго признака v_2^0 , третьего v_3^0 и так далее (v_4^0, \dots, v_n^0), т. е. аналогичным образом рассчитываются отношения

$$\frac{f_{ОИП}^1(v_1^0)}{f_{ЭП}^1(v_1^0)}, \frac{f_{ОИП}^2(v_2^0)}{f_{ЭП}^2(v_2^0)}, \frac{f_{ОВП}^1(v_1^0)}{f_{ЭП}^1(v_1^0)}, \frac{f_{ОВП}^2(v_2^0)}{f_{ЭП}^2(v_2^0)}. \quad (16)$$

Если после перебора всех признаков не удастся отнести субъект к тому или иному профилю с данным уровнем надежности (рассматриваемое отношение не выходит за пределы требуемых рубежей), то имеющиеся результаты обследования не позволяют принять решение с выбранным уровнем надежности. В этом случае можно понизить уровень надежности или обратиться за дополнительной информацией, связанной с получением новых признаков.

При отсутствии дополнительной информации для минимизации вероятности ошибки целесообразно принимать решение на основе принципа максимального правдоподобия [10].

В силу специального вида плотностей часто широко используются логарифмы встречаемых выше отношений, тогда все вычисления сводятся к последовательному сложению.

Отметим, что в данной работе задача классификации рассматривается в классической постановке [10], не учитывающей априорные вероятности классов. При наличии такой дополнительной априорной информации можно рассмотреть Байесовскую постановку и обобщить результаты на этот случай.

3. Градация признаков

При использовании любых количественных методов для отбора приходится прибегать к квантованию признака, так как часто не удастся достаточно точно измерить то или другое свойство человека, определяющее его индивидуальные психофизические способности. В таких случаях количество градаций зависит от способности дифференцировать данный признак. Если признак измеряется достаточно точно (например, время реакции), число градаций можно объективизировать. В рассматриваемом варианте, когда строятся одномерные распределения суммы интегральных признаков для психомоторного, кондиционного и координационного компонентов психофизических способностей, число градаций зависит от количества лиц в выборочной совокупности. Если число лиц невелико, квантование признака на диапазоны обеспечивается исходя из особенностей получающихся гистограмм, а также путем расчета средних арифметических величин и среднеквадратического отклонения (по правилу трех сигм) [7]. На основании исследований установлено, что в этих случаях достаточно от двух до четырех диапазонов. Когда распределения имеют сложную форму, диапазоны градаций являются неодинаковыми. В варианте ранжирования сумма интегральных баллов признаков переводится в три диапазона качественных критериев (низкий, средний и высокий уровень для каждого компонента) путем расчета средних арифметических величин и стандартного среднеквадратического отклонения [11].

4. Выбор порога

В последовательной статистической процедуре отношения правдоподобия сравниваются с порогами, зависящими от вероятностей ошибки $\alpha_{ОИП}$, $\alpha_{ОВП}$, $\beta_{ЭП}$, $\beta_{ОВП}$, $\gamma_{ЭП}$, $\gamma_{ОИП}$, что позволяет выбирать нужный порог, основываясь на сложившейся конъюнктуре [1].

Необходимость выбора небольшого числа лиц из больших контингентов делает возможным определить указанные единицы порядка 0,01 или даже 0,001. С другой стороны, при ограниченном количестве лиц естественно выбрать значения 0,05 или даже 0,10.

Если окажется, что ошибка пропустить хорошего специалиста и, наоборот, ошибка принять малопригодного неравноценны, то имеется возможность учесть это, выбирая разные вероятности $\alpha_{ОИП}$, $\alpha_{ОВП}$, $\beta_{ЭП}$, $\beta_{ОВП}$, $\gamma_{ЭП}$, $\gamma_{ОИП}$. В этой связи выбор порогов является весьма гибким и учитывает реальную обстановку, а также цену возможных ошибок.

Предлагаемый алгоритм имеет определенные преимущества перед другими способами определения профессиональной готовности и пригодности благодаря своей вычислительной простоте и удобству, а также эффективности. Данное обстоятельство связано с тем, что известные математические способы, используемые для целей определения профессиональной готовности и пригодности, как правило, предполагают нормальное распределение признаков, что в действительности не имеет места. Эффективность же предлагаемого алгоритма не зависит от вида распределений, а в случае независимости признаков, по-видимому, является и оптимальным методом разделения обследуемых субъектов на три военно-профессиональных профиля.

5. Пример применения алгоритма

Схема экспертной информационно-диагностической системы (ЭИДС) отражает последовательность оценки и прогнозирования психофизической готовности и пригодности субъектов из обследуемых контингентов по профилям военно-профессиональной деятельности (рис. 2).



Рис. 2. Система оценки психофизической готовности и пригодности к военно-профессиональной деятельности

Проверка эффективности применения предложенного алгоритма оценки и прогнозирования психофизической готовности и пригодности к конкретному профилю военно-профессиональной деятельности проводилась при обследовании суворовцев ($n = 80$) Минского суворовского военного училища (МСВУ) и кадетов ($n = 60$) Минского городского кадетского училища (МинГКУ). В каждом экстремальном, операторско-инженерном и общевоинском профиле имеются степени [2, 3]. Чем выше степень в конкретном профиле, тем выше уровень психофизической готовности и пригодности обследуемых кандидатов к определенной военно-профессиональной деятельности. На рис. 3 показаны результаты использования предложенного алгоритма.

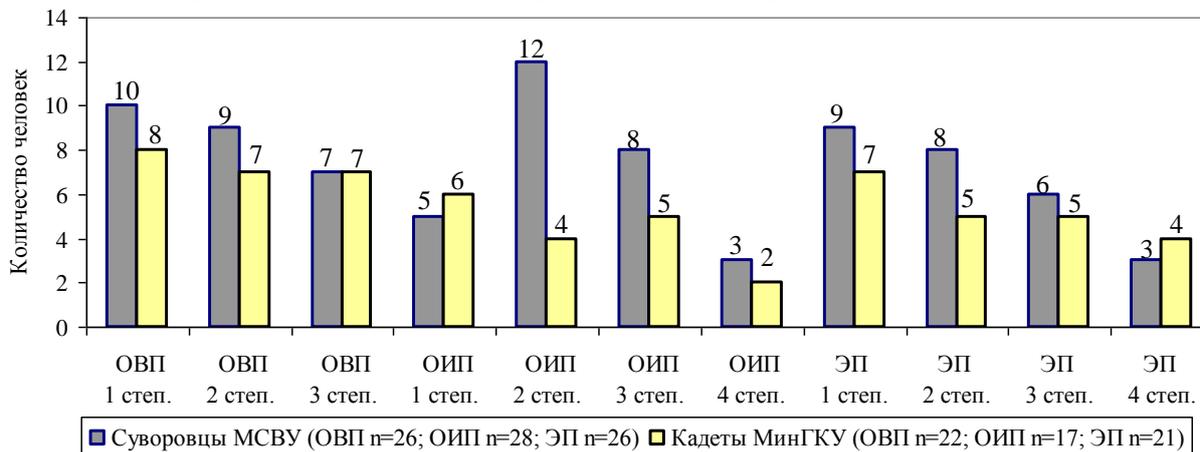


Рис. 3. Распределение учащихся по профилям военно-профессиональной деятельности

Распределение учащихся по степеням пригодности к профилям военно-профессиональной деятельности позволит проводить на начальном этапе становления военного специалиста целенаправленную подготовку с ориентацией на повышение профессионально важных качеств, способствующих формированию надежной психофизической готовности к продолжению службы в Вооруженных Силах Республики Беларусь.

Заключение

Рассмотренный алгоритм принятия решений является определением последовательного критерия отношения вероятностей в вариантах трех простых гипотез применительно к задаче последовательного анализа признаков.

Применяемый в алгоритме подход позволяет на научной основе оценивать и прогнозировать состояние готовности и пригодности учащейся молодежи к получению специального образования по ряду военных специальностей. При использовании методики диагностики психофизических характеристик кандидатов, обоснованных качественных и количественных критериев оценки результатов появляется возможность относить обследуемые субъекты с заданной малой вероятностью ошибки к конкретным степеням пригодности по профилям военно-профессиональной деятельности.

Предложенная схема экспертной информационно-диагностической системы будет способствовать совершенствованию учебно-воспитательного процесса в военных учебных заведениях при решении вопросов специализированного отбора контингента для определенного вида военно-профессиональной деятельности на начальном этапе профессионального становления.

Список литературы

1. Артемьева, Е.Ю. Вероятностные методы в психологии / Е.Ю. Артемьева. – М. : МГУ, 1985. – 206 с.
2. Михута, И.Ю. Диагностика психофизической готовности и пригодности суворовцев к военно-профессиональной деятельности / И.Ю. Михута // Вести БГПУ. – 2012. – № 2. – С. 12–16.

3. Михута, И.Ю. Структурно-содержательные модели специалистов военного профиля / И.Ю. Михута // Педагогіка, психологія та медико-біологічні проблеми фізичного виховання і спорту. – 2012. – № 4. – С. 34–38.
4. Бодров, В.А. Психология профессиональной пригодности : учебное пособие для вузов / В.А. Бодров. – М. : ПЕР СЭ, 2001. – 511 с.
5. Апчел, В.Я. Основные принципы и современные направления оценки профессиональной пригодности военных специалистов / В.Я. Апчел // Вестник Российской военно-медицинской академии: клинические исследования. – 2007. – № 5. – С. 60–64.
6. Большев, Л.Н. Таблицы математической статистики / Л.Н. Большев. – М. : Наука, 1983. – 416 с.
7. Голев, С.В. Математические методы в психологии / С.В. Голев. – Харьков : ИПИС. ХГУ, 2008. – 113 с.
8. Вентцель, Е.С. Теория вероятностей и математическая статистика : учебник / Е.С. Вентцель. – Минск : Выш. шк., 1999. – 576 с.
9. Гмурман, В.Е. Теория вероятностей и математическая статистика / В.Е. Гмурман. – Минск : Выш. шк., 2003. – 479 с.
10. Вальд, А. Последовательный анализ / А. Вальд. – М. : Наука, 1960. – 328 с.
11. Суходольский, Г.В. Математические методы в психологии / Г.В. Суходольский. – Харьков : Гуманитарный центр, 2004. – 284 с.

Поступила 12.07.12

*Белорусский государственный
педагогический университет им. М. Танка,
Минск, ул. Советская, 18
e-mail: igor_michuta@mail.ru*

I.Y. Mikhuta

**ALGORITHM OF THE ASSESSMENT AND FORECASTING
OF PSYCHOPHYSICAL READINESS AND SUITABILITY
OF STUDYING YOUTH TO MILITARY PROFESSIONAL ACTIVITY**

A statistical algorithm for the assessment and forecasting of psychophysical readiness and suitability of studying youth to military professional activity is proposed. The algorithm maps matrices of psychophysical criteria of readiness and suitability of subjects to concrete profiles (extreme, operator and engineering and general) of military professional activity.

ПАРАЛЛЕЛЬНЫЕ ВЫЧИСЛЕНИЯ

УДК 681.3.06:519

Н.С. Коваленко¹, В.Н. Венгеро², В.М. Метельский³МИНИМИЗАЦИЯ ЧИСЛА ОБРАБАТЫВАЮЩИХ УСТРОЙСТВ
ПРИ РАСПРЕДЕЛЕННОЙ ОРГАНИЗАЦИИ ВЫЧИСЛЕНИЙ

Рассматривается математическая модель организации неоднородных распределенных конкурирующих процессов при условии асинхронного режима их взаимодействия для систем с конвейерной организацией распределенных вычислений (СКОРВ). Предлагаются полиномиальные алгоритмы нахождения минимального числа обрабатывающих устройств (ОУ), которые обеспечивают выполнение заданных объемов вычислений за директивное или минимальное время.

Введение

Одной из центральных проблем, возникающих при исследовании и создании математических моделей и методов анализа и организации параллельных процессов, разработке программно-аппаратного обеспечения в СКОРВ, является нахождение условий оптимальной реализации заданных объемов вычислений [1]. Решение этой проблемы приводит к сложным комбинаторно-оптимизационным задачам, связанным с вычислением различных характеристик СКОРВ. Среди них важное место занимают задачи расчета минимального числа ОУ (процессоров и функциональных устройств в вычислительных системах, станков для обработки деталей, роботов, операторов конвейерных линий и др.), обеспечивающих выполнение заданных объемов вычислений (или работ) за минимальное либо заданное (директивное) время. Заметим, что именно эти задачи с практической точки зрения представляют значительный интерес. Методы решения такого рода задач достаточно проработаны в классической теории расписаний [2]. Однако открытым для исследований остается ряд прикладных задач, связанных с оптимальной организацией распределенных вычислений в СКОРВ.

Следует отметить, что для класса однородных сосредоточенных конкурирующих процессов при макроконвейерной либо конвейерной реализации вычислений задачи нахождения минимального числа ОУ решены в [3, 4]. До настоящего времени остаются нерешенными задачи расчета оптимального числа ОУ, обеспечивающих директивное либо минимальное время выполнения заданных объемов вычислений в условиях распределенной обработки неоднородных конкурирующих процессов [5]. В статье предлагаются алгоритмы для решения таких задач.

1. Основные понятия и математическая модель задачи

Как и в [3–6], под *процессом* будем понимать последовательность блоков (команд, процедур) Q_1, Q_2, \dots, Q_s , для выполнения которых используется множество ОУ. При этом процесс называется *распределенным*, если все блоки или часть из них обрабатываются разными ОУ. Для ускорения выполнения процессы могут обрабатываться параллельно.

Понятие *ресурса* применяется для обозначения любых объектов вычислительной системы, которые могут быть использованы процессами для своего выполнения. *Рееंतरабельные* (многократно применяемые) ресурсы характеризуются возможностью одновременного использования несколькими вычислительными процессами. Для СКОРВ характерной является ситуация, когда одну и ту же последовательность блоков или ее часть необходимо выполнять многократно несколькими ОУ; такую последовательность будем называть *программным ресурсом*, а соответствующие процессы – *конкурирующими*.

Математическая модель СКОРВ включает: p ($p \geq 2$) – число ОУ; s ($s \geq 2$) – число блоков линейно структурированного программного ресурса $PR = (Q_1, Q_2, \dots, Q_s)$; n ($n \geq 2$) –

число распределенных относительно PR конкурирующих процессов; $T = [t_{ij}]$ – матрицу времен выполнения j -х блоков i -ми конкурирующими процессами, $i = \overline{1, n}$; $j = \overline{1, s}$.

Так же, как в [3–6], будем считать, что взаимодействие процессов, ОУ и блоков линейно структурированного программного ресурса подчинено следующим условиям:

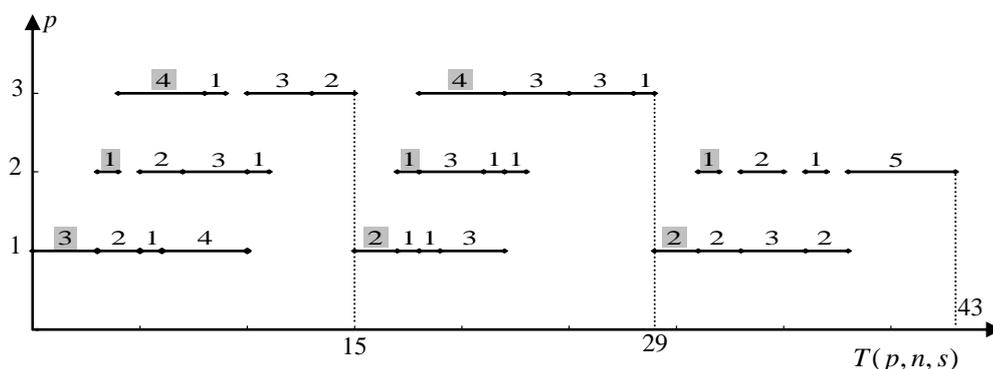
- 1) ни один из блоков PR не может обрабатываться одновременно более чем одним ОУ;
- 2) ни один из ОУ не может обрабатывать одновременно более одного блока;
- 3) обработка каждого блока осуществляется без прерываний;
- 4) распределение блоков программного ресурса по ОУ для каждого из процессов осуществляется циклически по правилу: блок с номером $j = kp + i$ ($j = \overline{1, s}$; $i = \overline{1, p}$; $k \geq 0$) распределяется на ОУ с номером i ;

5) отсутствуют простои ОУ при условии готовности блоков, а также невыполнение блоков при наличии ОУ.

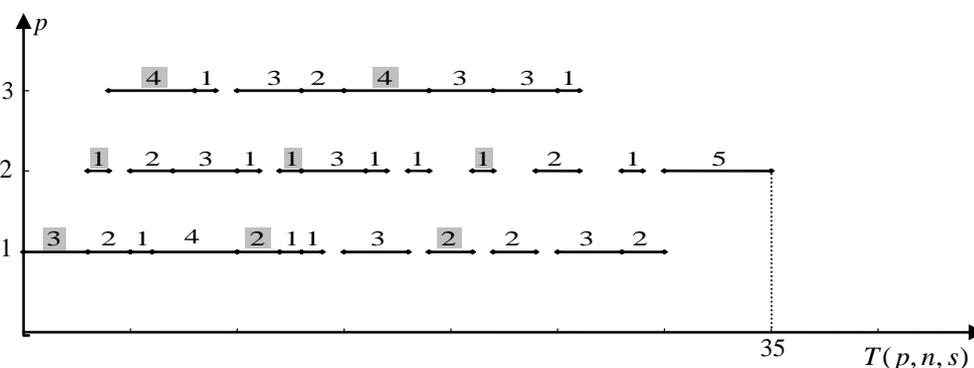
Условия 1–5 определяют *асинхронный режим* взаимодействия ОУ, процессов и блоков, который предполагает отсутствие простоев ОУ при условии готовности блоков, а также невыполнение блоков при наличии свободных ОУ.

СКОРВ называется *неоднородной*, если времена выполнения блоков PR зависят от объемов обрабатываемых данных и/или их структуры, т. е. являются разными для разных процессов.

На рис. 1 приведен пример распределенной СКОРВ для асинхронного режима взаимодействия процессов, ОУ и блоков, а также ее отображение с помощью линейных диаграмм Ганта.



а)



б)

Рис. 1. Асинхронный режим: а) несомещенная диаграмма Ганта; б) совмещенная

Пусть имеется система из n ($n \geq 2$) неоднородных конкурирующих процессов, готовых к выполнению в начальный момент времени и пронумерованных как 1, 2, ..., n . При этом предпо-

лагается, что все n процессов используют только одну копию программного ресурса. Пусть далее $T = [t_{ij}]$ – вещественная $(n \times s)$ -матрица, элемент t_{ij} которой определяет время выполнения j -го блока для i -го процесса, где $i = \overline{1, n}$; $j = \overline{1, s}$; d – заданное (директивное) время выполнения всех конкурирующих процессов.

Общее время выполнения всех конкурирующих процессов $T(p, n, s)$ при фиксированных n, s и заданном структурировании программного ресурса, определяемом $(n \times s)$ -матрицей $T = [t_{ij}]$, будет существенно зависеть от количества имеющихся ОУ [3, 4].

Задача состоит в том, чтобы при заданных n, s, d и матрице $T = [t_{ij}]$ (где $i = \overline{1, n}$; $j = \overline{1, s}$) найти оптимальное число ОУ, обеспечивающих директивное либо минимальное время выполнения неоднородных конкурирующих процессов.

Для решения поставленной задачи кроме введенных выше параметров математической модели p, n, s, d необходимы $\Gamma^{(q)}$ – двухмерный массив переменной длины, составленный специальным образом из элементов матрицы $T = [t_{ij}]$ (см., например, [6]), q ($q \in N$) – порядковый номер результирующей матрицы времен выполнения блоков (двухмерного массива переменной длины $\Gamma^{(q)}$), а также приведенные ниже определение и теорема.

О п р е д е л е н и е. Число ОУ в СКОРВ будем называть достаточным и обозначать p_* при заданных n, s , если $p_* = s$, ограниченным, если $p < s$, и избыточным, если $p > s$.

Ниже под критическим путем сетевого вершинно-взвешенного графа понимается путь с наибольшей суммой весов его вершин. Алгоритм его нахождения для линейных графов общеизвестен.

Если $t_{ij} = 0$, то это означает, что все соответствующие вершины графа равны нулю и не влияют на значение критического пути (приводятся для заполнения матрицы).

Пусть $T_*(p_*, n, s)$ – минимальное общее время выполнения множества конкурирующих процессов при достаточном числе ОУ, равном p_* ; $T(p, n, s)$ – минимальное общее время при ограниченном или избыточном числе ОУ, равном p .

Теорема. При заданных $n, s, T = [t_{ij}]$, $i = \overline{1, n}$, $j = \overline{1, s}$, в случаях достаточного ($p_* = s$) и ограниченного ($p < s$) числа ОУ в СКОРВ имеет место соотношение $T_*(p_*, n, s) \leq T(p, n, s)$.

Доказательство следует из того, что при достаточном числе p ОУ минимальное общее время $T_*(p_*, n, s)$ вычисляется как длина критического пути в сетевом вершинно-взвешенном графе, определяемом $(n \times s)$ -матрицей $T = [t_{ij}]$. При ограниченном числе $p < s$ ($s = kp + r$, $1 \leq r \leq p$) ОУ минимальное общее время $T(p < s, n, s)$, как показано в [6], определяется длиной критического пути в сетевом вершинно-взвешенном графе, задаваемом $(k+1)n \times (k+1)p$ -матрицей специального вида

$$\begin{bmatrix} \Gamma_1 & \Gamma_2 & \Gamma_3 & \cdots & \Gamma_k & \Gamma_{k+1} \\ \Gamma_2 & \Gamma_3 & \Gamma_4 & \cdots & \Gamma_{k+1} & Z_0 \\ \Gamma_3 & \Gamma_4 & \Gamma_5 & \cdots & Z_0 & Z_0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \Gamma_k & \Gamma_{k+1} & Z_0 & \cdots & Z_0 & Z_0 \\ \Gamma_{k+1} & Z_0 & Z_0 & \cdots & Z_0 & Z_0 \end{bmatrix}. \quad (1)$$

В матрице (1) $(n \times p)$ -матрицы Γ_l ($l = 1, \dots, k+1$) имеют вид

$$\begin{bmatrix} t_{1,(l-1)p+1} & t_{1,(l-1)p+2} & \cdots & t_{1,lp} \\ t_{2,(l-1)p+1} & t_{2,(l-1)p+2} & \cdots & t_{2,lp} \\ \vdots & \vdots & \ddots & \vdots \\ t_{n,(l-1)p+1} & t_{n,(l-1)p+2} & \cdots & t_{n,lp} \end{bmatrix}, \quad (2)$$

а $(n \times p)$ -матрицы Z_0 являются нулевыми. При этом в матрице Γ_{k+1} вида (2) при $r < p$ последние $p - r$ столбцов состоят из нулей.

Ввиду того что исходная $(n \times s)$ -матрица $T = [t_{ij}]$ входит в качестве подматрицы в первую строку матрицы (1), сетевой вершинно-взвешенный граф, определяемый $(n \times s)$ -матрицей $T = [t_{ij}]$, будет подграфом этого графа, определяемого матрицей (1). Так как длина критического пути в сетевом подграфе данного графа не может превышать такую длину в самом сетевом графе, то справедливо неравенство $T_*(p_*, n, s) \leq T(p, n, s)$, что и требовалось доказать.

Следует заметить, что в случае $p < s$ согласно принятой модели [6] достаточно использовать только s ОУ, а остальные $p - s$ не будут задействованы.

Данная теорема является отправной точкой для построения метода нахождения оптимального числа ОУ, обеспечивающих директивное и минимальное время выполнения неоднородных конкурирующих процессов при распределенной обработке в условиях асинхронного режима их взаимодействия в случаях достаточного ($p_* = s$) и ограниченного ($p < s$) чисел ОУ СКОРВ.

Пример 1. На рис. 1 представлены несовмещенная и совмещенная линейные диаграммы Ганта, с помощью которых отражено выполнение $n = 4$ неоднородных распределенных конкурирующих процессов в СКОРВ с $p = 3$ ОУ при $s = 8$ блоках структурированного программного ресурса. Длительности выполнения каждого из блоков указаны на диаграммах, причем для первого процесса они составляют (3, 1, 4, 2, 1, 4, 2, 1), второго – (2, 2, 1, 1, 3, 3, 2, 2), третьего – (1, 3, 3, 1, 1, 3, 3, 1), четвертого – (4, 1, 2, 3, 1, 1, 2, 5). Для наглядности длительности блоков первого процесса выделены. Общее время выполнения $T(p, n, s)$ неоднородных распределенных конкурирующих процессов на несовмещенной диаграмме равно 43, а на совмещенной – 35.

На рис. 2 для данного примера представлен сетевой граф, соответствующий совмещенной линейной диаграмме Ганта (см. рис. 1, б). Вершины данного графа содержат соответственно номера процессов (первая цифра), блоков (вторая цифра) и ОУ (третья цифра).

2. Оптимизация числа обрабатываемых устройств при директивном времени

Предлагается следующий алгоритм минимизации числа ОУ при директивном времени.

Входные данные: p – заданное (исходное) число ОУ ($p \geq 2$); n – число конкурирующих неоднородных распределенных процессов ($n \geq 2$); s – число блоков линейно-структурированного программного ресурса ($s \geq 2$); $\Gamma^{(0)}$ – двухмерный массив, содержащий элементы исходной $(n \times s)$ -матрицы $T = [t_{ij}]$; d – заданное (директивное) время выполнения конкурирующих процессов.

Выходные данные: \bar{p}_0 – минимальное число ОУ, обеспечивающих выполнение конкурирующих процессов за директивное время; $\Gamma^{(q)}$ – двухмерный массив, содержащий результирующую матрицу времен выполнения блоков программного ресурса; q – порядковый номер двухмерного массива, содержащего результирующую матрицу времен выполнения блоков программного ресурса.

Случай 1. Если $d < T_*(p_*, n, s)$, то полагаем $\bar{p}_0 = 0$, т. е. директивное время выполнения конкурирующих процессов d не может быть реализовано в заданных условиях ни для какого числа ОУ.

Случай 2. Пусть $d \geq T_*(p_*, n, s)$ и число ОУ в СКОРВ является ограниченным, т. е. $p < s$. Тогда между d , $T_*(p_*, n, s)$ и $T(p, n, s)$ возможны следующие соотношения:

– если $T_*(p_*, n, s) \leq d = T(p, n, s)$ или $d > T(p, n, s)$, то для нахождения \bar{p}_0 используется метод дихотомии, применяемый к отрезку $[2, p]$;

– если $T_*(p_*, n, s) \leq d < T(p, n, s)$, то нахождение \bar{p}_0 осуществляется методом дихотомии, применяемым к отрезку $[p, p_*]$.

Случай 3. Пусть $p \geq s$. Тогда нахождение \bar{p}_0 осуществляется методом дихотомии, применяемым к отрезку $[2, p_*]$.

Нетрудно подсчитать, что сложность алгоритма нахождения оптимального числа ОУ \bar{p}_0 , базирующегося на предложенном методе, составляет в худшем случае величину $\hat{I}((k+1)^2 np \log_2 p)$ операций. Действительно, поиск с использованием алгоритма деления заданного отрезка пополам составляет не более $\log_2 p$ операций, подсчет критического пути в сетевом графе с линейным числом дуг и числом вершин $(k+1)n \times (k+1)p$ составляет не более $(k+1)n \times (k+1)p$ операций. Таким образом, получаем указанную выше оценку сложности алгоритма нахождения оптимального числа ОУ при заданном директивном времени.

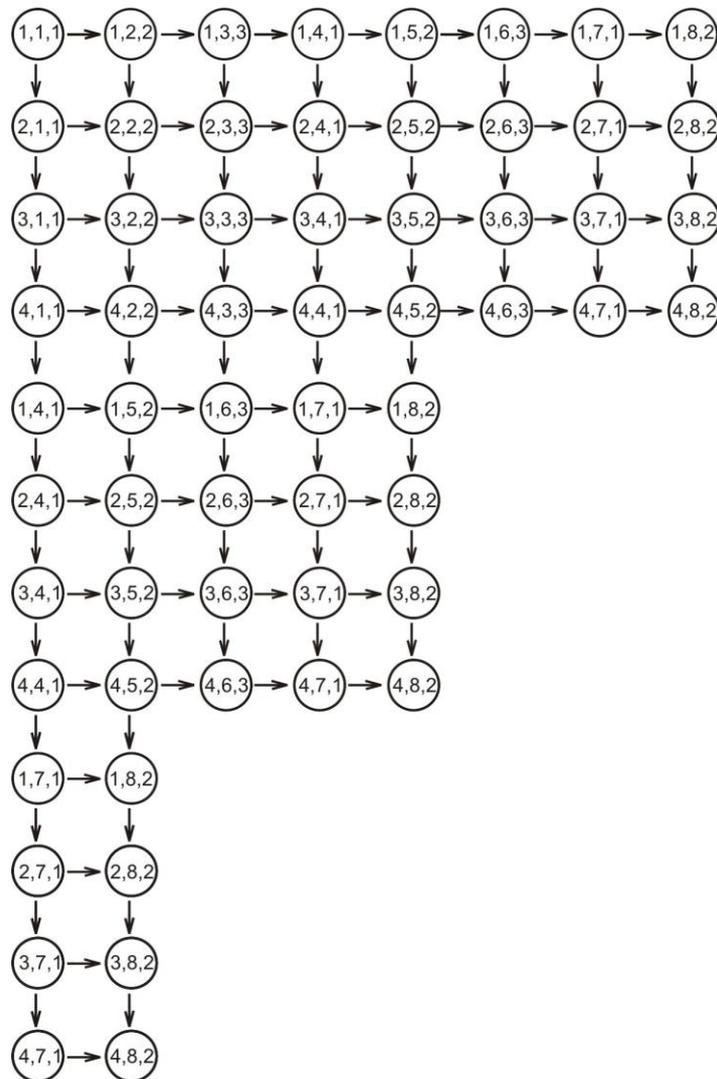


Рис. 2. Сетевой граф для совмещенной линейной диаграммы Ганта

На рис. 3 приведена графическая интерпретация зависимости величины $T(p, n, s)$ от числа p ОУ, а также указаны величины d , $T_*(p_*, n, s)$, \bar{p}_0 и p_* . Видно, что величина \bar{p}_0 определяется либо как точка пересечения прямой d с дискретной линией, определяющей зависимость $T(p, n, s)$ от p , либо как ближайшая точка, которая находится ниже прямой d .

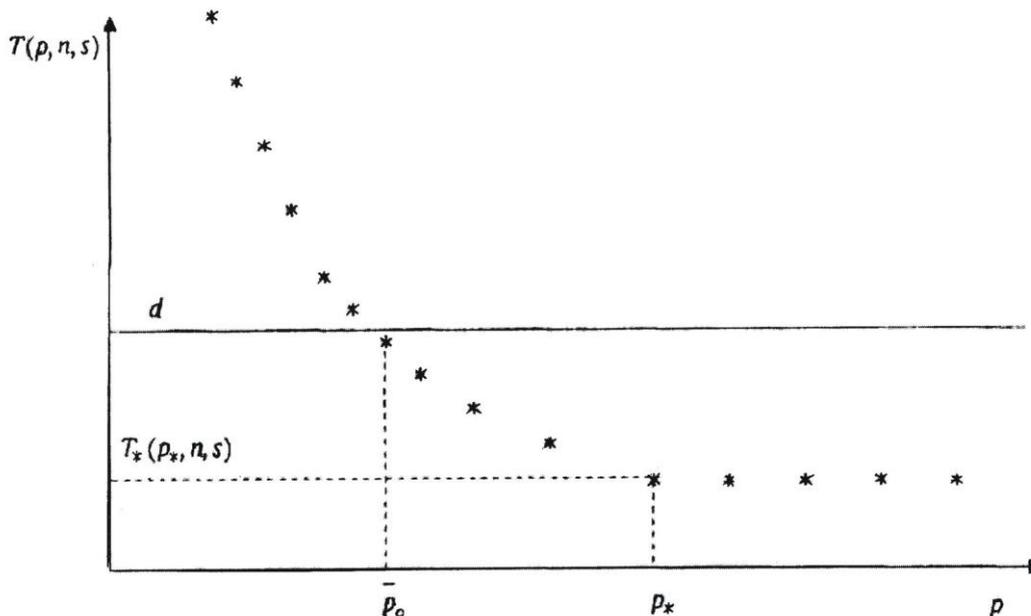


Рис. 3. Графическая интерпретация зависимости величины $T(p, n, s)$ от числа p ОУ

Пример 2. Пусть $d = 50$ — заданное (директивное) время выполнения конкурирующих процессов, $p = 3$ — число ОУ в СКОРВ, $n = 3$ — число конкурирующих неоднородных процессов, $s = 9$ — число блоков линейно-структурированного программного ресурса, а исходная матрица времен выполнения блоков программного ресурса имеет вид

$$\Gamma^{(0)} = \begin{bmatrix} 5 & 2 & 4 & 6 & 1 & 7 & 4 & 8 & 6 \\ 1 & 7 & 3 & 2 & 5 & 8 & 3 & 1 & 2 \\ 6 & 5 & 2 & 4 & 9 & 3 & 6 & 1 & 3 \end{bmatrix}.$$

В данном случае достаточное число ОУ $p_* = 9$.

1. По исходной матрице $\Gamma^{(0)}$ строим сетевой вершинно-взвешенный граф и находим величину $T_*(p_* = 9, n, s) = 48$ (длину критического пути) [6]. Далее по исходным значениям p, n, s и $\Gamma^{(0)}$ строим результирующую матрицу вида (1), дополненную матрицей вида (2):

$$\left[\begin{array}{c} \begin{bmatrix} 5 & 2 & 4 \\ 1 & 7 & 3 \\ 6 & 5 & 2 \end{bmatrix} \\ \begin{bmatrix} 6 & 1 & 7 \\ 2 & 5 & 8 \\ 4 & 9 & 3 \end{bmatrix} \\ \begin{bmatrix} 4 & 8 & 6 \\ 3 & 1 & 2 \\ 6 & 1 & 3 \end{bmatrix} \\ \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \end{array} \right].$$

С помощью полученной матрицы вычисляем величину $T(p = 3, n, s) = 53$. Учитывая, что $T_*(p_* = 9, n, s) = 48 \leq d = 50 < T(p = 3, n, s) = 53$, рассмотрим отрезок $[3, 9]$.

2. Методом деления отрезка $[3, 9]$ пополам находим $p_1 = 6$ и строим по заданным n, s и полученному значению $p_1 = 6$ результирующую матрицу $\Gamma^{(1)}$ вида (1), (2):

$$\Gamma^{(1)} = \begin{bmatrix} \begin{bmatrix} 5 & 2 & 4 & 6 & 1 & 7 \\ 1 & 7 & 3 & 2 & 5 & 8 \\ 6 & 5 & 2 & 4 & 9 & 3 \\ 4 & 8 & 6 & 0 & 0 & 0 \\ 3 & 1 & 2 & 0 & 0 & 0 \\ 6 & 1 & 3 & 0 & 0 & 0 \end{bmatrix} & \begin{bmatrix} 4 & 8 & 6 & 0 & 0 & 0 \\ 3 & 1 & 2 & 0 & 0 & 0 \\ 6 & 1 & 3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \end{bmatrix}.$$

С помощью полученной матрицы $\Gamma^{(1)}$ вычисляем величину $T(p_1 = 6, n, s) = 48$. Так как $T(p_1 = 6, n, s) = 48 \leq d = 50$, то рассматриваем отрезок $[3, 6]$.

3. Методом деления отрезка $[3, 6]$ пополам находим $p_2 = 4$, причем в качестве p_2 берем величину, которая является наименьшим целым полусуммы чисел 3 и 6. Далее по заданным n, s и полученному значению $p_2 = 4$ строим результирующую матрицу $\Gamma^{(2)}$:

$$\tilde{A}^{(2)} = \begin{bmatrix} \begin{bmatrix} 5 & 2 & 4 & 6 \\ 1 & 7 & 3 & 2 \\ 6 & 5 & 2 & 4 \\ 1 & 7 & 4 & 8 \\ 5 & 8 & 3 & 1 \\ 9 & 3 & 6 & 1 \end{bmatrix} & \begin{bmatrix} 1 & 7 & 4 & 8 \\ 5 & 8 & 3 & 1 \\ 9 & 3 & 6 & 1 \\ 6 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 \\ 3 & 0 & 0 & 0 \end{bmatrix} & \begin{bmatrix} 6 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 \\ 3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \end{bmatrix}.$$

С помощью полученной матрицы $\Gamma^{(2)}$ вычисляем величину $T(p_2 = 4, n, s) = 49$. Таким образом, директивное время выполнения $n=3$ процессов реализуется при $p_2 = 4$, так как $d = 50 > T(p_2 = 4, n, s) = 49$, и не реализуется при $p = 3$, так как $d = 50 < T(p = 3, n, s) = 53$. Следовательно, $\bar{p}_0 = 4$.

3. Оптимизация числа обрабатываемых устройств при минимальном времени

Предлагается следующий алгоритм минимизации числа ОУ при минимальном времени.

Входные данные: p – заданное (исходное) число ОУ ($p \geq 2$); n – число конкурирующих неоднородных распределенных процессов ($n \geq 2$); s – число блоков линейно-структурированного программного ресурса ($2 \leq s \leq p$); $\Gamma^{(0)}$ – двумерный массив, содержащий элементы исходной $(n \times s)$ -матрицы $T = [t_{ij}]$; q – порядковый номер двумерного массива, содержащего результирующую матрицу времен выполнения блоков программного ресурса.

Выходные данные: $T_*(p_*, n, s)$ – минимально возможное время выполнения программного ресурса конкурирующими процессами; \bar{p}_0 – минимальное число ОУ, обеспечивающих минимально возможное время выполнения программного ресурса конкурирующими процессами.

Случай 1 ($s \leq p$)

1. Находим достаточное число ОУ ($p=s$) и вычисляем минимальное общее время (т. е. $T_*(p_*, n, s)$) при достаточном числе ОУ.

2. Уменьшаем число ОУ p на единицу (начиная с $p=s$) до тех пор, пока оно не станет меньше двух или пока время $T(p, n, s)$, полученное при новом значении p , не станет больше $T_*(p_*, n, s)$.

Случай 2 ($s > p$)

1. Находим достаточное число ОУ ($p=s$) и вычисляем минимальное общее время при достаточном числе ОУ, т. е. $T_*(p_*, n, s)$.

2. Вычисляем $T(p, n, s)$ – время выполнения блоков программного ресурса при заданном числе ОУ p :

– если $T(p, n, s) = T_*(p_*, n, s)$, то уменьшаем исходное (а не достаточное) число ОУ на единицу до тех пор, пока p не станет меньше двух или пока время $T(p, n, s)$ не станет больше чем $T_*(p_*, n, s)$;

– если $T(p, n, s) > T_*(p_*, n, s)$, то увеличиваем число ОУ на единицу до тех пор, пока время $T(p, n, s)$ не станет равным $T_*(p_*, n, s)$.

Пример 3. Пусть $p=9$ – число ОУ в СКОРВ, $n=3$ – число конкурирующих неоднородных процессов, $s=9$ – число блоков линейно-структурированного программного ресурса. Пусть также задана исходная матрица времен выполнения блоков программного ресурса:

$$\Gamma^{(0)} = \begin{bmatrix} 5 & 2 & 4 & 6 & 1 & 7 & 4 & 8 & 6 \\ 1 & 7 & 3 & 2 & 5 & 8 & 3 & 1 & 2 \\ 6 & 5 & 2 & 4 & 9 & 3 & 6 & 1 & 3 \end{bmatrix}.$$

С помощью приведенного выше метода находим минимальное число ОУ \bar{p}_0 , обеспечивающих минимально возможное время выполнения конкурирующих процессов:

1. Определяем достаточное число ОУ, обрабатывающих блоки программного ресурса: $p=9, q=0$.

2. Вычисляем длину критического пути в ориентированном вершинно-взвешенном графе, определяемом матрицей $\Gamma^{(0)}$: $T_*(p=9, n, s) = 48$.

3. Уменьшаем число ОУ на единицу: $p=8, p > 2, q=1$.

4. Строим результирующую матрицу $\Gamma^{(1)}$ размерности 6×16 по заданным $n=3, s=9$ и новому значению $p=8$:

$$\tilde{A}^{(1)} = \left[\begin{array}{c} \left[\begin{array}{cccccccc} 5 & 2 & 4 & 6 & 1 & 7 & 4 & 8 \end{array} \right] \\ \left[\begin{array}{cccccccc} 1 & 7 & 3 & 2 & 5 & 8 & 3 & 1 \end{array} \right] \\ \left[\begin{array}{cccccccc} 6 & 5 & 2 & 4 & 9 & 3 & 6 & 1 \end{array} \right] \\ \left[\begin{array}{cccccccc} 6 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array} \right] \\ \left[\begin{array}{cccccccc} 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array} \right] \\ \left[\begin{array}{cccccccc} 3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array} \right] \end{array} \right] \cdot \left[\begin{array}{cccccccc} \left[\begin{array}{cccccccc} 6 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array} \right] \\ \left[\begin{array}{cccccccc} 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array} \right] \\ \left[\begin{array}{cccccccc} 3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array} \right] \\ \left[\begin{array}{cccccccc} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array} \right] \\ \left[\begin{array}{cccccccc} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array} \right] \\ \left[\begin{array}{cccccccc} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array} \right] \end{array} \right].$$

5. Вычисляем длину критического пути в орграфе, определяемом матрицей $\Gamma^{(1)}$: $T(p=8, n, s) = 48, T(p=8, n, s) = T_*(p=9, n, s) = 48$.

6. Уменьшаем число ОУ на единицу: $p=7, p > 2, q=2$.

7. Строим результирующую матрицу $\Gamma^{(2)}$ размерности 6×14 по заданным $n=3, s=9$ и новому значению $p=7$:

$$\tilde{A}^{(2)} = \left[\begin{array}{c} \left[\begin{array}{cccccc} 5 & 2 & 4 & 6 & 1 & 7 & 4 \\ 1 & 7 & 3 & 2 & 5 & 8 & 3 \\ 6 & 5 & 2 & 4 & 9 & 3 & 6 \end{array} \right] \\ \left[\begin{array}{cccccc} 8 & 6 & 0 & 0 & 0 & 0 & 0 \\ 1 & 2 & 0 & 0 & 0 & 0 & 0 \\ 1 & 3 & 0 & 0 & 0 & 0 & 0 \end{array} \right] \\ \left[\begin{array}{cccccc} 8 & 6 & 0 & 0 & 0 & 0 & 0 \\ 1 & 2 & 0 & 0 & 0 & 0 & 0 \\ 1 & 3 & 0 & 0 & 0 & 0 & 0 \end{array} \right] \end{array} \right].$$

8. Вычисляем длину критического пути в орграфе, определяемом матрицей $\Gamma^{(2)}$:
 $T(p=7, n, s) = 48$, $T(p=7, n, s) = T_*(p=9, n, s) = 48$.

9. Уменьшаем число ОУ на единицу: $p=6$, $p > 2$, $q=3$.

10. Строим результирующую матрицу $\Gamma^{(3)}$ размерности 6×12 по заданным $n=3$, $s=9$ и новому значению $p=6$:

$$\tilde{A}^{(3)} = \left[\begin{array}{c} \left[\begin{array}{cccccc} 5 & 2 & 4 & 6 & 1 & 7 \\ 1 & 7 & 3 & 2 & 5 & 8 \\ 6 & 5 & 2 & 4 & 9 & 3 \end{array} \right] \\ \left[\begin{array}{cccccc} 4 & 8 & 6 & 0 & 0 & 0 \\ 3 & 1 & 2 & 0 & 0 & 0 \\ 6 & 1 & 3 & 0 & 0 & 0 \end{array} \right] \\ \left[\begin{array}{cccccc} 4 & 8 & 6 & 0 & 0 & 0 \\ 3 & 1 & 2 & 0 & 0 & 0 \\ 6 & 1 & 3 & 0 & 0 & 0 \end{array} \right] \end{array} \right].$$

11. Вычисляем длину критического пути в орграфе, определяемом матрицей $\Gamma^{(3)}$:
 $T(p=6, n, s) = 48$, $T(p=6, n, s) = T_*(p=9, n, s) = 48$.

12. Уменьшаем число ОУ на единицу: $p=5$, $p > 2$, $q=4$.

13. Строим результирующую матрицу $\Gamma^{(4)}$ размерности 6×10 по заданным $n=3$, $s=9$ и новому значению $p=5$:

$$\tilde{A}^{(4)} = \left[\begin{array}{c} \left[\begin{array}{ccccc} 5 & 2 & 4 & 6 & 1 \\ 1 & 7 & 3 & 2 & 5 \\ 6 & 5 & 2 & 4 & 9 \end{array} \right] \\ \left[\begin{array}{ccccc} 7 & 4 & 8 & 6 & 0 \\ 8 & 3 & 1 & 2 & 0 \\ 3 & 6 & 1 & 3 & 0 \end{array} \right] \\ \left[\begin{array}{ccccc} 7 & 4 & 8 & 6 & 0 \\ 8 & 3 & 1 & 2 & 0 \\ 3 & 6 & 1 & 3 & 0 \end{array} \right] \end{array} \right].$$

14. Вычисляем длину критического пути в орграфе, определяемом матрицей $\Gamma^{(4)}$:
 $T(p=5, n, s) = 48$, $T(p=5, n, s) = T_*(p=9, n, s) = 48$.

15. Уменьшаем число ОУ на единицу: $p=4$, $p > 2$, $q=5$.

16. Строим результирующую матрицу $\Gamma^{(5)}$ размерности 9×12 по заданным $n=3$, $s=9$ и новому значению $p=4$:

$$\tilde{A}^{(5)} = \left[\begin{array}{c} \left[\begin{array}{cccc} 5 & 2 & 4 & 6 \\ 1 & 7 & 3 & 2 \\ 6 & 5 & 2 & 4 \end{array} \right] \\ \left[\begin{array}{cccc} 1 & 7 & 4 & 8 \\ 5 & 8 & 3 & 1 \\ 9 & 3 & 6 & 1 \end{array} \right] \\ \left[\begin{array}{cccc} 6 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 \\ 3 & 0 & 0 & 0 \end{array} \right] \\ \left[\begin{array}{cccc} 1 & 7 & 4 & 8 \\ 5 & 8 & 3 & 1 \\ 9 & 3 & 6 & 1 \end{array} \right] \\ \left[\begin{array}{cccc} 6 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 \\ 3 & 0 & 0 & 0 \end{array} \right] \\ \left[\begin{array}{cccc} 1 & 7 & 4 & 8 \\ 5 & 8 & 3 & 1 \\ 9 & 3 & 6 & 1 \end{array} \right] \\ \left[\begin{array}{cccc} 6 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 \\ 3 & 0 & 0 & 0 \end{array} \right] \end{array} \right].$$

17. Вычисляем длину критического пути в орграфе, определяемом матрицей $\Gamma^{(5)}$: $T(p=4, n, s) = 49$, $T(p=4, n, s) > T_*(p=9, n, s)$.

18. $\bar{p}_0 = 5$.

Таким образом, минимально возможное время выполнения программного ресурса конкурирующими процессами $T_*(p_*, n, s) = 48$, а минимальное число ОУ, обеспечивающих выполнение программного ресурса за это время, равно 5.

Пример 4. Пусть $p=3$ – число ОУ в СКОРВ, $n=4$ – число конкурирующих неоднородных процессов, $s=9$ – число блоков линейно-структурированного программного ресурса. Пусть также задана исходная матрица времен выполнения блоков программного ресурса:

$$\tilde{A}^{(0)} = \begin{bmatrix} 2 & 5 & 1 & 4 & 7 & 2 & 3 & 8 & 1 \\ 6 & 3 & 2 & 8 & 5 & 1 & 7 & 2 & 3 \\ 3 & 7 & 5 & 2 & 6 & 3 & 1 & 4 & 6 \\ 1 & 6 & 2 & 3 & 1 & 5 & 6 & 1 & 2 \end{bmatrix}.$$

Результирующая матрица $\Gamma^{(1)}$ времен выполнения блоков программного ресурса в этом случае будет иметь вид

$$\tilde{A}^{(1)} = \begin{bmatrix} \begin{bmatrix} 2 & 5 & 1 \\ 6 & 3 & 2 \\ 3 & 7 & 5 \\ 1 & 6 & 2 \end{bmatrix} & \begin{bmatrix} 4 & 7 & 2 \\ 8 & 5 & 1 \\ 2 & 6 & 3 \\ 3 & 1 & 5 \end{bmatrix} & \begin{bmatrix} 3 & 8 & 1 \\ 7 & 2 & 3 \\ 1 & 4 & 6 \\ 6 & 1 & 2 \end{bmatrix} \\ \begin{bmatrix} 4 & 7 & 2 \\ 8 & 5 & 1 \\ 2 & 6 & 3 \\ 3 & 1 & 5 \end{bmatrix} & \begin{bmatrix} 3 & 8 & 1 \\ 7 & 2 & 3 \\ 1 & 4 & 6 \\ 6 & 1 & 2 \end{bmatrix} & \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \\ \begin{bmatrix} 3 & 8 & 1 \\ 7 & 2 & 3 \\ 1 & 4 & 6 \\ 6 & 1 & 2 \end{bmatrix} & \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} & \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \end{bmatrix}.$$

Находим минимальное число \bar{p}_0 ОУ, обеспечивающих выполнение множества конкурирующих процессов за минимальное время. Из теоремы следует, что это время не может превышать минимального общего времени выполнения конкурирующих процессов при достаточном числе ОУ:

1. $q = 1$.

2. Находим длину критического пути в орграфе, определяемом исходной матрицей $\Gamma^{(0)}$: $T_*(p=9, n, s) = 46$.

3. Вычисляем длину критического пути в орграфе, определяемом результирующей матрицей $\Gamma^{(1)}$: $T(p=3, n, s) = 3$, $T(p=3, n, s) > T_*(p=9, n, s)$.

4. Увеличиваем на единицу число ОУ, обрабатывающих программный ресурс: $p=4$, $q=2$.

5. Строим новую результирующую матрицу $\Gamma^{(2)}$ размерности 12×12 по заданным n, s и новому значению $p=4$:

$$\tilde{A}^{(2)} = \begin{bmatrix} \begin{bmatrix} 2 & 5 & 1 & 4 \\ 6 & 3 & 2 & 8 \\ 3 & 7 & 5 & 2 \\ 1 & 6 & 2 & 3 \end{bmatrix} & \begin{bmatrix} 7 & 2 & 3 & 8 \\ 5 & 1 & 7 & 2 \\ 6 & 3 & 1 & 4 \\ 1 & 5 & 6 & 1 \end{bmatrix} & \begin{bmatrix} 1 & 0 & 0 & 0 \\ 3 & 0 & 0 & 0 \\ 6 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 \end{bmatrix} \\ \begin{bmatrix} 7 & 2 & 3 & 8 \\ 5 & 1 & 7 & 2 \\ 6 & 3 & 1 & 4 \\ 1 & 5 & 6 & 1 \end{bmatrix} & \begin{bmatrix} 1 & 0 & 0 & 0 \\ 3 & 0 & 0 & 0 \\ 6 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 \end{bmatrix} & \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \\ \begin{bmatrix} 1 & 0 & 0 & 0 \\ 3 & 0 & 0 & 0 \\ 6 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 \end{bmatrix} & \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} & \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \end{bmatrix}.$$

6. Находим длину критического пути в орграфе, определяемом матрицей $\Gamma^{(2)}$:
 $T(p=4, n, s) = 52, T(p=4, n, s) > T_*(p=9, n, s)$.

7. Увеличиваем число ОУ, обрабатывающих программный ресурс, на единицу: $p=5, q=3$.

8. Строим новую результирующую матрицу $\Gamma^{(3)}$ размерности 8×10 по заданным n, s и новому значению $p=5$:

$$\tilde{A}^{(3)} = \begin{bmatrix} \begin{bmatrix} 2 & 5 & 1 & 4 & 7 \\ 6 & 3 & 2 & 8 & 5 \\ 3 & 7 & 5 & 2 & 6 \\ 1 & 6 & 2 & 3 & 1 \end{bmatrix} & \begin{bmatrix} 2 & 3 & 3 & 1 & 0 \\ 1 & 7 & 7 & 3 & 0 \\ 3 & 1 & 1 & 6 & 0 \\ 5 & 6 & 6 & 2 & 0 \end{bmatrix} \\ \begin{bmatrix} 2 & 3 & 3 & 1 & 0 \\ 1 & 7 & 7 & 3 & 0 \\ 3 & 1 & 1 & 6 & 0 \\ 5 & 6 & 6 & 2 & 0 \end{bmatrix} & \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \end{bmatrix}.$$

9. Находим длину критического пути в орграфе, определяемом матрицей $\Gamma^{(3)}$:
 $T(p=5, n, s) = 48, T(p=5, n, s) > T_*(p=9, n, s)$.

10. $p_0 = 5$. Выход.

Итак, минимальное число ОУ, обеспечивающих выполнение программного ресурса за минимальное время $T_*(p=9, n, s) = 49$, равно 5. Выигрыш по времени в результате применения алгоритма составляет 16 единиц (или 24,6 %).

Заключение

В статье исследован ряд прикладных задач, связанных с оптимальной организацией распределенных вычислений в системах с конвейерной организацией распределенных вычислений. Для данных систем построена математическая модель организации неоднородных распределенных конкурирующих процессов при условии асинхронного режима их взаимодействия. Для нахождения минимального числа обрабатывающих устройств при выполнении заданных объемов вычислений за директивное либо минимальное время предложены алгоритмы, которые имеют полиномиальную сложность, что особенно важно для их практических приложений. Работа алгоритмов наглядно подтверждена конкретными примерами, которые показывают их эффективность и возможность простого и удобного использования на практике.

Список литературы

1. Абламейко, С.В. Принципы построения суперкомпьютеров семейства СКИФ и их реализация / С.В. Абламейко [и др.] // Информатика. – 2004. – № 1. – С. 89–106.
2. Танаев, В.С. Теория расписаний. Групповые технологии / В.С. Танаев, М.Я. Ковалев, Я.М. Шафранский. – Минск : Ин-т техн. кибернетики НАН Беларуси, 1998. – 290 с.

3. Капитонова, Ю.В. Метод структурирования программных ресурсов при организации параллельных вычислений / Ю.В. Капитонова, Н.С. Коваленко // Докл. IV Всесоюз. симпоз. «Системное и теоретическое программирование». – Кишинев : Штиинца, 1983. – С. 183–185.

4. Овсеец, М.И. Минимизация числа обрабатываемых устройств при реализации однородных конкурирующих процессов / М.И. Овсеец // Доклады АН БССР. – 1985. – № 12. – С. 1082–1085.

5. Иванников, В.П. О минимальном времени реализации конкурирующих процессов в синхронных режимах / В.П. Иванников, Н.С. Коваленко, В.М. Метельский // Программирование. – 2000. – № 5. – С. 268–274.

6. Коваленко, Н.С. О времени реализации конкурирующих процессов при распределенной обработке / Н.С. Коваленко, В.М. Метельский // Кибернетика и системный анализ. – 1996. – № 1. – С. 54–64.

Поступила 07.02.2013

¹Белорусский государственный
экономический университет,
Минск, пр. Партизанский, 26
e-mail: kovalenkons@rambler.ru

²Объединенный институт проблем
информатики НАН Беларуси,
Минск, ул. Сурганова, 6
e-mail: vengerov@basnet.by

³Белорусский государственный университет
информатики и радиоэлектроники,
Минск, ул. П. Бровки, 6
e-mail: metmargen@mail.ru

N.S. Kovalenko, V.N. Vengherov, V.M. Metelyskij

MINIMIZING THE NUMBER OF PROCESSING UNITS IN DISTRIBUTED COMPUTATIONS

Polynomial algorithms for finding the minimum number of processing units in the problem of scheduling heterogeneous competing processes on parallel processors are developed. The algorithms ensure minimum completion time of all the processes and their completion by the specified deadline.

УДК 004.9; 004.4; 004.056

А.А. Коляда, А.Ф. Чернявский

ИНТЕГРАЛЬНО-ХАРАКТЕРИСТИЧЕСКАЯ БАЗА МОДУЛЯРНЫХ СИСТЕМ СЧИСЛЕНИЯ

Рассматривается проблематика создания интегрально-характеристической базы модулярных систем счисления, определенных на диапазонах неотрицательных целых чисел. Для решения поставленной задачи применяется аппарат интервально-модулярных форм целых чисел, ключевую роль в котором выполняют интервально-индексные характеристики – интервальный индекс и главный интервальный индекс. Приоритетные позиции данных характеристик обусловлены их существенными преимуществами над известными интегральными характеристиками модулярного кода при оптимизации алгоритмов немодулярных операций.

Введение

Постоянный интерес к модулярной вычислительной технологии как к уникальному средству распараллеливания вычислений стимулирует разработки по созданию и оптимизации классов модулярных вычислительных структур, которые ориентированы на обеспечение принципиально нового уровня производительности и контроля достоверности расчетов на сложных математических моделях, таких, в частности, как множества комплексных и гиперкомплексных чисел, полиномов и т. п. [1–5]. Для модулярных вычислительных структур, определенных на диапазонах многомерного типа, вещественные модулярные системы счисления (МСС) выполняют роль систем нижнего уровня. Поэтому разработка и оптимизация алгоритмов компьютерной арифметики МСС данного класса имеют основополагающее значение.

Модулярный код явно не содержит информации о величине отвечающего ему элемента рабочего диапазона. Поэтому при выполнении в МСС операций, тем или иным образом связанных с некоторыми характеристиками местоположения целых чисел (ЦЧ) в диапазоне или за его пределами, приходится использовать формы представления ЦЧ (через цифры модулярного кода), позволяющие получить искомые характеристики. В отличие от модулярных операций (сложения, вычитания, умножения без контроля переполнения), реализуемых поразрядно (параллельно), операции указанного типа в модулярной арифметике квалифицируются как немодулярные. Базовые формы ЦЧ для таких операций включают одну или более интегральных характеристик модулярного кода (ИХМК) – числовых характеристик, которые рассчитываются по части или всем цифрам данного кода.

Вполне понятно, что сложность вычисления применяемых ИХМК в конечном счете определяет эффективность созданной на их основе модулярной арифметики. В свете сказанного в общей проблематике разработки модулярной арифметики ключевая роль принадлежит исследованиям по оптимизации интегрально-характеристической базы МСС, и в первую очередь МСС с вещественными (целочисленными) диапазонами. В настоящей статье излагаются теоретические основы универсальной технологии расчета ИХМК, составляющие так называемый аппарат интервально-модулярных форм (ИМФ). В рамках развиваемого подхода в качестве вспомогательного инструментария используется также и ранговая форма ЦЧ, которая наряду с полиадической формой традиционно применяется для синтеза немодулярных процедур [6–8]. Преимущества, обеспечиваемые ИМФ, обусловлены модульностью базовой ИХМК – интервального индекса (ИИ). Благодаря данному свойству ИИ приведение ЦЧ к остаткам по модулю с помощью ИМФ существенно упрощается. При этом интервально-индексные характеристики связаны с другими ИХМК тривиальными соотношениями. Это относится к характеристикам не только одного и того же, но и разных порядков, а также к ИХМК элементов симметричных диапазонов. Реализация преимуществ ИМФ и связанных с ними интервально-индексных характеристик применительно к проблеме построения интегрально-характеристической базы модулярной арифметики составляет главное содержание представляемых исследований.

инные совокупности представителей классов вычетов. В компьютерных приложениях их роль выполняют множества \mathbf{Z}_{M_k} или $\mathbf{Z}_{M_k}^-$. С учетом сказанного в первом случае модулярное кодирование определяется как отображение $v_{MCC}: \mathbf{Z}_{M_k} \rightarrow \mathbf{Z}_{m_1} \times \mathbf{Z}_{m_2} \times \dots \times \mathbf{Z}_{m_k}$, которое каждому $X \in \mathbf{Z}_{M_k}$ ставит в соответствие код $(\chi_1, \chi_2, \dots, \chi_k)$. Декодирующее отображение $v_{MCC}^{-1}: \mathbf{Z}_{m_1} \times \mathbf{Z}_{m_2} \times \dots \times \mathbf{Z}_{m_k} \rightarrow \mathbf{Z}_{M_k}$, базирующееся на (2), действует согласно правилу

$$X = \left| \sum_{i=1}^k M_{i,k} \chi_{i,k} \right| \quad (\chi_{i,k} \equiv M_{i,k}^{-1} \chi_i \pmod{m_i}) \quad (3)$$

Непосредственное применение выражения (3) в качестве базовой формы ЦЧ при синтезе немодульных процедур практически невозможно из-за сложности прямой компьютерной реализации, особенно в случае больших M_k . Вместе с тем из (3) с помощью специальных ИХМК могут быть получены параллельные формы ЦЧ, обладающие весьма хорошими реализационными свойствами.

Пусть по набору модулей m_1, m_2, \dots, m_l ($2 \leq l \leq k$) числу $X \in \mathbf{Z}$ отвечает модулярный код $(\chi_1, \chi_2, \dots, \chi_l)$. Согласно КТО (теореме 1) по аналогии с (3) можно записать:

$$|X|_{M_l} \equiv \left| \sum_{i=1}^l M_{i,l} \left| M_{i,l}^{-1} \chi_i \right|_{m_i} \right|_{M_l}. \quad (4)$$

Из (4) следует, что разность $|X|_{M_l} - \sum_{i=1}^l M_{i,l} \left| M_{i,l}^{-1} \chi_i \right|_{m_i}$ кратна константе M_l . Таким образом, справедливо равенство

$$\begin{aligned} |X|_{M_l} &= \sum_{i=1}^l M_{i,l} \left| M_{i,l}^{-1} \chi_i \right|_{m_i} - M_l \rho_l(X) = \\ &= \sum_{i=1}^l M_{i,l} \chi_{i,l} - M_l \rho_l(X) \quad (\chi_{i,l} \equiv \left| M_{i,l}^{-1} \chi_i \right|_{m_i}), \end{aligned} \quad (5)$$

где $\rho_l(X)$ – некоторое ЦЧ. При любом $X \in \mathbf{Z}$ числу $|X|_{M_l} \in \mathbf{Z}_{M_l}$ соответствует единственное значение ИХМК $\rho_l(X) = \rho_l(\chi_1, \chi_2, \dots, \chi_l)$.

Определение 1. ИХМК $\rho_l(X)$ называется рангом числа $|X|_{M_l}$ в МСС с основаниями m_1, m_2, \dots, m_l и диапазоном \mathbf{Z}_{M_l} или ранговой характеристикой l -го порядка, а выражение (5) – ранговой формой ЦЧ того же порядка.

Согласно КТО, применяемой к системе сравнений

$$\begin{cases} X \equiv \chi_1 \pmod{m_1}, \\ X \equiv \chi_2 \pmod{m_2}, \\ \dots \dots \dots \\ X \equiv \chi_{l-1} \pmod{m_{l-1}}, \end{cases}$$

имеем $X \equiv (\sum_{i=1}^{l-1} M_{i,l-1} |M_{i,l-1}^{-1} \chi_i|_{m_i}) \pmod{M_{l-1}}$. Следовательно, для X существует единственное ЦЧ (обозначим его через $I_l(X)$), такое, что

$$\begin{aligned} X &= \sum_{i=1}^{l-1} M_{i,l-1} |M_{i,l-1}^{-1} \chi_i|_{m_i} + M_{l-1} I_l(X) = \\ &= \sum_{i=1}^{l-1} M_{i,l-1} \chi_{i,l-1} + M_{l-1} I_l(X) \quad (\chi_{i,l-1} = |M_{i,l-1}^{-1} \chi_i|_{m_i}). \end{aligned} \quad (6)$$

Определение 2. ИХМК $I_l(X)$ называется ИИ ЦЧ X относительно модулей m_1, m_2, \dots, m_l , а выражение вида (6) – ИМФ ЦЧ.

При $l=k$ для введенных ИХМК $\rho_l(X)$ и $I_l(X)$ будем использовать также обозначения $\rho(X)$ и $I(X)$.

Определение 3. Компоненты $\hat{I}_l(X) = |I_l(X)|_{m_l}$ и $J_l(X) = \lfloor I_l(X)/m_l \rfloor$ представления ИИ $I_l(X)$ вида

$$I_l(X) = \hat{I}_l(X) + m_l J_l(X) \quad (7)$$

назовем соответственно компьютерным и главным ИИ ЦЧ X относительно модулей m_1, m_2, \dots, m_l .

С помощью ранговой и интервально-индексной характеристик, а также связанных с ними форм ЦЧ могут быть реализованы все немодульные операции. Наряду с ранговой и интервально-индексной версиями модулярной арифметики на практике часто применяется версия, использующая полиадическую форму ЦЧ [6, 11, 12], которая имеет вид

$$|X|_{M_l} = \sum_{i=1}^l M_{i-1} x_i \quad (X \in \mathbf{Z}; M_0 = 1; x_i \in \mathbf{Z}_{m_i}). \quad (8)$$

В модулярной арифметике, построенной на основе (8), коэффициенты x_1, x_2, \dots, x_l выполняют роль ИХМК.

Определение 4. Систему счисления (СС), в которой ЦЧ из диапазона \mathbf{Z}_{M_l} представляется в форме (8), называют обобщенной позиционной СС, полиадической СС или СС со смешанным основанием. При этом для $|X|_{M_l}$ употребляется запись $|X|_{M_l} = \langle x_l x_{l-1} \dots x_1 \rangle_{m_1, m_2, \dots, m_l}$ или $|X|_{M_l} = \langle x_l x_{l-1} \dots x_1 \rangle$.

Определение 5. Величину $N_l(X) = \lfloor X/M_l \rfloor$ назовем интервальным номером ЦЧ X относительно модулей m_1, m_2, \dots, m_l ($l \geq 1$).

3. Вычислительная структура интегральных характеристик кода избыточных МСС с целочисленными диапазонами

Для введенных ИХМК верны приводимые ниже утверждения.

Теорема 2. Максимальное значение $\rho_{l, \max}$ ранга $\rho_l(X) = \rho_l(|X|_{M_l})$ числа $|X|_{M_l} = (\chi_1, \chi_2, \dots, \chi_l)$ ($X \in \mathbf{Z}; l > 1$) в МСС с основаниями m_1, m_2, \dots, m_l не превышает $l-1$, т. е.

$$\rho_l(X) \leq \rho_{l, \max} = \max \{ \rho_l(A) \mid A \in \mathbf{Z}_{M_l} \} \leq l-1. \quad (9)$$

Доказательство. Согласно (5)

$$|X|_{M_l} = \sum_{i=1}^l M_{i,l} - M_l \rho_l(X) \quad (X \in \mathbf{Z}; \chi_{i,l} = |M_{i,l}^{-1} \chi_i|_{m_i}). \quad (10)$$

Деление (10) M_l с последующим переходом в обеих частях полученного равенства к антье $\lfloor |X|_{M_l} / M_l \rfloor = 0$ дает

$$\rho_l(X) = \left\lfloor \sum_{i=1}^l \frac{\chi_{i,l}}{m_i} \right\rfloor. \quad (11)$$

Пусть $m_{\max,l} = \max\{m_1, m_2, \dots, m_l\}$. Так как $m_{\max,l} > l$ при любом $l > 1$, из (11) с учетом взаимной однозначности отображений $\chi_i \rightarrow \chi_{i,l}$ на \mathbf{Z}_{m_i} ($i = \overline{1, l}$) для $\rho_l(X)$ вытекает оценка

$$\rho_l(X) \leq \left\lfloor \sum_{i=1}^l \frac{m_i - 1}{m_i} \right\rfloor = l + \left\lfloor -\sum_{i=1}^l \frac{1}{m_i} \right\rfloor = l - \left\lceil \sum_{i=1}^l \frac{1}{m_i} \right\rceil \leq l - \left\lceil \sum_{i=1}^l \frac{1}{m_{\max,l}} \right\rceil = l - \left\lceil \frac{l}{m_{\max,l}} \right\rceil = l - 1. \blacksquare$$

Теорема 3 (о ранге числа). В МСС с основаниями $m_1, m_2, \dots, m_{l-1}, m_l \geq l - 2$ ($l > 1$) и диапазоном \mathbf{Z}_{M_l} ранг $\rho_l(X) = \rho_l(|X|_{M_l})$ числа $|X|_{M_l} = (\chi_1, \chi_2, \dots, \chi_l)$ ($X \in \mathbf{Z}$) представим в виде

$$\rho_l(X) = \hat{\rho}_l(X) + \Theta_l(X), \quad (12)$$

где

$$\hat{\rho}_l(X) = \left\lfloor m_l^{-1} \sum_{i=1}^l R_{i,l}(\chi_i) \right\rfloor; \quad (13)$$

$$R_{i,l}(\chi_i) = \left\lfloor \frac{m_l}{m_i} |M_{i,l}^{-1} \chi_i|_{m_i} \right\rfloor = \lfloor -m_i^{-1} |M_{i,l-1}^{-1} \chi_i|_{m_i} \rfloor_{m_i} \quad (i = \overline{1, l-1});$$

$$R_{l,l}(\chi_l) = \left\lfloor \frac{\chi_l}{M_{l-1}} \right\rfloor_{m_l}; \quad (14)$$

$\Theta_l(X)$ – двузначная величина, принимающая значения 0 или 1.

Доказательство. Применяя лемму Евклида [9], можно записать

$$\begin{aligned} m_l \chi_{i,l} &= |m_l \chi_{i,l}|_{m_i} + m_i \left\lfloor \frac{m_l}{m_i} \chi_{i,l} \right\rfloor = |m_l |M_{i,l}^{-1} \chi_i|_{m_i}|_{m_i} + m_i \left\lfloor \frac{m_l}{m_i} |M_{i,l}^{-1} \chi_i|_{m_i} \right\rfloor = \\ &= \chi_{i,l-1} + m_i R_{i,l}(\chi_i) \quad (i = \overline{1, l-1}). \end{aligned} \quad (15)$$

Фигурирующая в (15) величина $R_{i,l}(\chi_i) = \left\lfloor \frac{m_l}{m_i} |M_{i,l}^{-1} \chi_i|_{m_i} \right\rfloor \in \mathbf{Z}_{m_i}$. Поэтому переход в (15) к остаткам по модулю m_l позволяет получить $R_{i,l}(\chi_i)$ в следующей эквивалентной форме:

$$R_{i,l}(\chi_i) = (m_l \chi_{i,l} - \chi_{i,l-1}) / m_i = \lfloor -m_i^{-1} |M_{i,l-1}^{-1} \chi_i|_{m_i} \rfloor_{m_l}.$$

Используя (15) и обозначения (14), выполним над ранговой формой ЦЧ $|X|_{M_l}$ (см. (5)) преобразование

$$\begin{aligned} |X|_{M_l} &= \sum_{i=1}^l M_{i,l} \chi_{i,l} - M_l \rho_l(X) = \sum_{i=1}^{l-1} M_{i,l-1} m_i \chi_{i,l} + M_{l,l} \chi_{l,l} - M_l \rho_l(X) = \\ &= \sum_{i=1}^{l-1} M_{i,l-1} \chi_{i,l-1} + M_{l-1} \left(\sum_{i=1}^l R_{i,l}(\chi_i) - m_l \rho_l(X) \right) = \sum_{i=1}^{l-1} M_{i,l-1} \chi_{i,l-1} - M_{l-1} \rho_{l-1}(X) + \\ &\quad + M_{l-1} \rho_{l-1}(X) + M_{l-1} \left(\sum_{i=1}^l R_{i,l}(\chi_i) - m_l \rho_l(X) \right). \end{aligned} \quad (16)$$

Для МСС с основаниями m_1, m_2, \dots, m_{l-1} и диапазоном $\mathbf{Z}_{M_{l-1}}$ аналог равенства (5) (ранговая форма ЦЧ $(l-1)$ -го порядка) имеет вид

$$|X|_{M_{l-1}} = \sum_{i=1}^{l-1} M_{i,l-1} \chi_{i,l-1} - M_{l-1} \rho_{l-1}(X). \quad (17)$$

Благодаря (17) из (16) вытекает соотношение

$$|X|_{M_l} = |X|_{M_{l-1}} + M_{l-1}(\rho_{l-1}(X) + \sum_{i=1}^l R_{i,l}(\chi_i) - m_l \rho_l(X)).$$

В соответствии с леммой Евклида отсюда следует, что

$$N_{l-1}(|X|_{M_{l-1}}) = \lfloor |X|_{M_l} / M_{l-1} \rfloor = \rho_{l-1}(X) + \sum_{i=1}^l R_{i,l}(\chi_i) - m_l \rho_l(X)$$

или

$$\rho_l(X) + (N_{l-1}(|X|_{M_l}) - \rho_{l-1}(X)) / m_l = m_l^{-1} \sum_{i=1}^l R_{i,l}(\chi_i). \quad (18)$$

Теперь для получения искомого результата – формулы (12) – достаточно в (18) перейти к антье, приняв при этом во внимание (13), а также обозначение

$$\Theta_l(X) = -\lfloor (N_{l-1}(|X|_{M_l}) - \rho_{l-1}(X)) / m_l \rfloor = \left\lfloor (\rho_{l-1}(X) - N_{l-1}(|X|_{M_l})) / m_l \right\rfloor. \quad (19)$$

Согласно теореме 2, применяемой к МСС с основаниями m_1, m_2, \dots, m_{l-1} , для ранговой характеристики $\rho_{l-1}(X)$ верна оценка $\rho_{l-1}(X) \leq l-2$. Поэтому благодаря условию $m_l > l-2$ ввиду

$$N_{l-1}(|X|_{M_l}) \in \mathbf{Z}_{m_l} \text{ справедливо неравенство } -1 < -\frac{m_l-1}{m_l} \leq (\rho_{l-1}(X) - N_{l-1}(|X|_{M_l})) / m_l \leq \frac{l-2}{m_l} \leq 1,$$

из которого вытекает двузначность величины (19): $\Theta_l(X) \in \{0, 1\}$. ■

Определение 6. Величину $\Theta_l(X)$, определяемую формулой (19), назовем минимальной ИХМК l -го порядка, отвечающей ЦЧ X в МСС с основаниями m_1, m_2, \dots, m_{l-1} и диапазоном \mathbf{Z}_{M_l} ($l > 1$).

Приведенное доказательство теоремы 3 позволяет сформулировать следующее утверждение.

Теорема 4. Минимальная ИХМК $\Theta_l(X)$, отвечающая произвольному ЦЧ X в МСС с основаниями m_1, m_2, \dots, m_{l-1} , $m_l \geq l-2$ ($l > 1$), является двузначной величиной: $\Theta_l(X) \in \{0, 1\}$.

Теорема 5 (об интервальном индексе ЦЧ). Для ИИ $I_l(X)$ произвольного элемента $X = (\chi_1, \chi_2, \dots, \chi_l)$ диапазона \mathbf{Z}_{M_l} МСС с основаниями m_1, m_2, \dots, m_{l-1} , $m_l \geq l-2$ ($l > 1$) справедлива формула

$$I_l(X) = \hat{I}_l(X) - m_l \Theta_l(X), \quad (20)$$

где

$$\hat{I}_l(X) = \left\lfloor \sum_{i=1}^l R_{i,l}(\chi_i) \right\rfloor_{m_l}; \quad (21)$$

вычеты $R_{i,l}(\chi_i)$ вычисляются по (14); $\Theta_l(X)$ – минимальная ИХМК l -го порядка вида (19) ($\Theta_l(X) \in \{0, 1\}$).

Доказательство. Используя (6) и (17), представим ЦЧ X следующим образом:

$$X = \sum_{i=1}^{l-1} M_{i,l-1} \chi_{i,l-1} + M_{l-1} I_l(X) = \sum_{i=1}^{l-1} M_{i,l-1} \chi_{i,l-1} - M_{l-1} \rho_{l-1}(X) + M_{l-1} I_l(X) = |X|_{M_{l-1}} + M_{l-1} (\rho_{l-1}(X) + I_l(X)).$$

Отсюда согласно лемме Евклида вытекает равенство $N_{l-1}(X) = \lfloor X / M_{l-1} \rfloor = \rho_{l-1}(X) + I_l(X)$, из которого находим

$$\lfloor I_l(X) / m_l \rfloor = \lfloor (N_{l-1}(X) - \rho_{l-1}(X)) / m_l \rfloor = -\lceil (\rho_{l-1}(X) - N_{l-1}(X)) / m_l \rceil. \quad (22)$$

Принимая во внимание (19) и определение (18), на основании (22) заключаем, что для главного ИИ $G_l(X)$ ЦЧ $X \in \mathbf{Z}_{M_l}$ в заданной МСС верна формула

$$G_l(X) = -\Theta_l(X). \quad (23)$$

Подстановка (23) в (7) доказывает справедливость (20).

Расчетное соотношение (21) для компьютерного ИИ $\hat{I}_l(X)$ ЦЧ X вытекает из (6):

$$\hat{I}_l(X) = |I_l(X)|_{m_l} = -\sum_{i=1}^{l-1} \frac{\chi_{i,l-1}}{m_i} + \frac{X}{M_{l-1}} \Big|_{m_l} = \left| \sum_{i=1}^{l-1} -m_i^{-1} |M_{i,l-1}^{-1} \chi_i|_{m_i} \Big|_{m_l} + \left| M_{l-1}^{-1} \chi_l \Big|_{m_l} \right|_{m_l} = \left| \sum_{i=1}^l R_{i,l}(\chi_i) \right|_{m_l}$$

(см. (14)). ■

Теорема 6 (об интервальном номере ЦЧ). Для интервального номера $N_l(X) = \lfloor X / M_l \rfloor$ произвольного ЦЧ X относительно $m_1, m_2, \dots, m_{l-1}, m_l \geq l-2$ имеет место равенство $N_l(X) = J_l(X) + \Theta_l(X)$, где $J_l(X)$ – главный ИИ числа X , а $\Theta_l(X)$ – отвечающая ему минимальная ИХМК l -го порядка ($\Theta_l(X) \in \{0, 1\}$).

Доказательство. Пусть числу X в МСС с основаниями m_1, m_2, \dots, m_l и диапазоном \mathbf{Z}_{M_l} отвечает модулярный код $(\chi_1, \chi_2, \dots, \chi_l)$. Используя (7) и (20), выполним над ИМФ ЦЧ X l -го порядка (см. (6)) преобразование

$$\begin{aligned} X &= \sum_{i=1}^{l-1} M_{i,l-1} \chi_{i,l-1} + M_{l-1} I_l(X) = \sum_{i=1}^{l-1} M_{i,l-1} \chi_{i,l-1} + M_{l-1} (\hat{I}_l(X) + m_l J_l(X)) = \\ &= \sum_{i=1}^{l-1} M_{i,l-1} \chi_{i,l-1} + M_{l-1} (\hat{I}_l(X) - m_l \Theta_l(X) + m_l \Theta_l(X) + m_l J_l(X)) = \\ &= \sum_{i=1}^{l-1} M_{i,l-1} \chi_{i,l-1} + M_{l-1} I_l(|X|_{M_l}) + M_l (J_l(X) + \Theta_l(X)) = |X|_{M_l} + M_l (J_l(X) + \Theta_l(X)). \end{aligned}$$

Таким образом, $X = |X|_{M_l} + M_l (J_l(X) + \Theta_l(X))$. Отсюда в соответствии с леммой Евклида, а также определением 5 и вытекает искомый результат: $\lfloor X / M_l \rfloor = N_l(X) = J_l(X) + \Theta_l(X)$. ■

Теорема 7 (о полиадическом коде числа). Пусть в МСС с основаниями m_1, m_2, \dots, m_k задан произвольный элемент $X = (\chi_1, \chi_2, \dots, \chi_k)$ диапазона \mathbf{Z}_{M_k} и пусть

$$L_l(X) = \sum_{i=1}^{l-1} M_{i,l-1} \chi_{i,l-1} + M_{l-1} \hat{I}_l(X) \quad (\chi_{i,l-1} = |M_{i,l-1}^{-1} \chi_i|_{m_i}; 2 \leq l \leq k), \quad (24)$$

$\hat{I}_l(X)$ определяется по (21) с использованием (14). Тогда для коэффициентов полиадической формы числа X (определение 4)

$$X = \sum_{i=1}^k M_{i-1} x_i \quad (M_0 = 1; x_i \in \mathbf{Z}_{m_i}) \quad (25)$$

верны формулы

$$x_1 = \chi_1, x_2 = \hat{I}_2(X), x_3 = \hat{x}_3, x_l = |\hat{x}_l + \Theta_{l-1}(X)|_{m_l} \quad (l = \overline{4, k}), \quad (26)$$

где

$$\hat{x}_l = |J_{l-1}(L_l(X))|_{m_l} \quad (l = \overline{3, k}); \quad (27)$$

$J_{l-1}(L_l(X))$ – главный ИИ ЦЧ $L_l(X)$ в МСС с основаниями m_1, m_2, \dots, m_{l-1} , вычисляемый по правилу

$$J_{l-1}(L_l(X)) = \hat{\rho}_{l-1}(X) + \hat{I}_l(X); \quad (28)$$

$$\hat{\rho}_{l-1}(X) = \left[m_{l-1}^{-1} \sum_{i=1}^{l-1} R_{i,l-1}(\chi_i) \right]; \quad (29)$$

вычеты $R_{i,l-1}(\chi_i)$ определяются по (14) с заменой l на $l-1$; $\Theta_{l-1}(X)$ – минимальная ИХМК $(l-1)$ -го порядка, которая при $m_{l-1} \geq l-3$ принимает значения 0 или 1 (см. теорему 4).

Доказательство. Из (25) следует, что при $l = \overline{1, k-1}$ выполняется равенство $|X|_{M_l} = \sum_{i=1}^l M_{i-1}x_i$, в результате чего

$$N_{l-1}(|X|_{M_l}) = \lfloor |X|_{M_l} / M_{l-1} \rfloor = x_l. \quad (30)$$

Согласно (6), (7) и (24)

$$X = \sum_{i=1}^{l-1} M_{l-1} \chi_{i,l-1} + M_{l-1} (\hat{I}_l(X) + m_l J_l(X)) = L_l(X) + M_l J_l(X).$$

Поэтому $|X|_{M_l} = |L_l(X)|_{M_l}$ и (30) можно записать в следующей эквивалентной форме:

$$x_l = N_{l-1}(|L_l(X)|_{M_l}). \quad (31)$$

Поскольку по лемме Евклида $L_l(X) = |L_l(X)|_{M_l} + M_l N_l(L_l(X))$, то $N_{l-1}(L_l(X)) = N_{l-1}(|L_l(X)|_{M_l}) + m_l N_l(L_l(X))$. Отсюда ввиду (31) заключаем, что

$$x_l = |N_{l-1}(L_l(X))|_{m_l}. \quad (32)$$

Применяя к интервальному номеру $N_l(L_l(X))$ теорему 6 из (32), находим

$$x_l = |J_{l-1}(L_l(X)) + \Theta_{l-1}(X)|_{m_l}. \quad (33)$$

С учетом обозначения (27) равенство (33) совпадает с искомым соотношением для x_l при $l = \overline{4, k}$ (см. (26)).

Как видно из (5), в одномодульном (вырожденном) случае (в случае $l = 1$) для любого ЦЧ X ранг $\rho_1(X) = 0$, вследствие чего и $\Theta_2(X) = 0$ (см. (19)). Поэтому при $l = 3$ (33) дает $x_3 = |J_2(L_3(X))|_{m_3}$. Искомые равенства для коэффициентов x_1 и x_2 вытекают соответственно из (30) и (32):

$$x_1 = N_0(|X|_{M_1}) = \lfloor |X|_{M_1} / M_0 \rfloor = |X|_{M_1} = \chi_1;$$

$$x_2 = |N_1(L_2(X))|_{m_2} \left\| \left[\frac{1}{m_1} (\chi_1 + m_1 \hat{I}_2(X)) \right] \right\|_{m_2} = \hat{I}_2(X).$$

Что касается расчетных соотношений (28), (29), то для их вывода воспользуемся процедурой сужения ИМФ ЦЧ. Необходимая процедура состоит в получении ИМФ $(l-1)$ -го порядка числа по его ИМФ l -го порядка. Осуществляемое сужение базируется на соотношении типа (15). Замена в (15) l на $l-1$ приводит к равенствам

$$m_{l-1} \chi_{i,l-1} = \chi_{i,l-2} + m_{l-1} R_{i,l-1}(\chi_i) \quad (i = \overline{1, l-2}). \quad (34)$$

Применяя (34), выполним над (24) преобразование

$$\begin{aligned} L_l(X) &= \sum_{i=1}^{l-2} M_{i,l-2} m_{l-1} \chi_{i,l-1} + M_{l-1,l-1} \chi_{l-1,l-1} + M_{l-1} \hat{I}_l(X) = \\ &= \sum_{i=1}^{l-2} M_{i,l-2} \chi_{i,l-2} + M_{l-2} \left(\sum_{i=1}^{l-1} R_{i,l-1}(\chi_i) + m_{l-1} \hat{I}_l(X) \right). \end{aligned} \quad (35)$$

В соответствии с (6) из (35) находим

$$I_{l-1}(L_l(X)) = \sum_{i=1}^{l-1} R_{i,l-1}(\chi_i) + m_{l-1} \hat{I}_l(X). \quad (36)$$

Деление (36) на m_{l-1} с последующим переходом в обеих частях полученного равенства к антье дает искомым результат – расчетные соотношения (28), (29).

4. Метод сужения ИМФ ЦЧ для определения минимальных ИХМК

Основой для расчета минимальных ИХМК по разработанной интервально-индексной технологии служат приводимые ниже теоремы, а также операция сужения ИМФ ЦЧ.

Теорема 8. Для минимальной ИХМК $\Theta_l(X)$, отвечающей числу $X \in Z$ в МСС с основаниями $m_1, m_2, \dots, m_{l-1}, m_l \geq l-2$ ($l > 1$), справедлива формула

$$\Theta_l(X) = 1 - \text{sn}(Z_l(X)), \quad (37)$$

где

$$Z_l(X) = L_l(X) - M_l; \quad (38)$$

ЦЧ $L_l(X)$ определяется соотношением (24) с использованием модулярного кода $(\chi_1, \chi_2, \dots, \chi_l)$ ($\chi_i = |X|_{m_i}$ ($i = \overline{1, l}$)) и интервально-индексной характеристики $\hat{I}_l(X)$ (см. (21)); через $\text{sn}(x)$ обозначается знаковая функция вида

$$\text{sn}(x) = \begin{cases} 0, & \text{если } x \geq 0; \\ 1, & \text{если } x < 0. \end{cases}$$

Доказательство. Подставляя (24) в (38) и применяя (6), а также теорему 5, получим

$$\begin{aligned} Z_l(X) &= \sum_{i=1}^{l-1} M_{i,l-1} \chi_{i,l-1} + M_{l-1} \hat{I}_l(X) - M_l = \sum_{i=1}^{l-1} M_{i,l-1} \chi_{i,l-1} + M_{l-1} (\hat{I}_l(X) - \\ &- m_l \Theta_l(X) + m_l \Theta_l(X)) - M_l = \sum_{i=1}^{l-1} M_{i,l-1} \chi_{i,l-1} + M_{l-1} I_l(|X|_{M_l}) + M_l (\Theta_l(X) - 1) = \\ &= |X|_{M_l} + M_l (\Theta_l(X) - 1). \end{aligned} \quad (39)$$

По теореме 4 $\Theta_l(X) \in \{0, 1\}$. При этом из (39) ввиду $0 \leq X|_{M_l} < M_l$ следует, что значению $\Theta_l(X)=0$ ИХМК $\Theta_l(X)$ соответствует $Z_l(X) < 0$, а значению $\Theta_l(X)=1 - Z_l(X) \geq 0$. Таким образом, формула (37) верна. ■

Теорема 9. Пусть число $X \in Z$ по набору модулей m_1, m_2, \dots, m_k отвечает модулярный код $(\chi_1, \chi_2, \dots, \chi_k)$ и пусть $J_l(X)$ – главный ИИ ЦЧ X относительно $m_1, m_2, \dots, m_{l-1}, m_l \geq l-2$ ($2 \leq l \leq k$). Знаки чисел X и $J_l(X)$ совпадают при $l=2$, а также при $l>2$, если $J_l(X) \neq -1$.

Доказательство. Согласно лемме Евклида и теореме 6 об интервальном номере ЦЧ выполняется равенство

$$X = |X|_{M_l} + M_l N_l(X) = |X|_{M_l} + M_l (J_l(X) + \Theta_l(X)), \quad (40)$$

где $\Theta_l(X)$ – минимальная ИХМК, отвечающая X в МСС с базисом $\{m_1, m_2, \dots, m_l\}$, причем $\Theta_l(X) \in \{0, 1\}$. Как отмечалось в ходе доказательства теоремы 7, ИХМК $\Theta_2(X) = 0$, поэтому из (40) следует, что $\text{sn}(X) = \text{sn}(J_2(X))$. При $l>2$ формирование $\text{sn}(X)$ числа X по его главному ИИ $J_l(X)$ невозможно только тогда, когда $J_l(X) = -1$. В этом случае возникает неопределенная ситуация, обусловленная тем, что знак $\text{sn}(N_l(X))$ интервального номера $N_l(X) = \Theta_l(X) - 1$ ввиду $\Theta_l(X) = 0$ или 1 неоднозначен: $\text{sn}(\Theta_l(X) - 1) \in \{0, 1\}$. Если $J_l(X) \neq -1$, то $\text{sn}(X) = \text{sn}(N_l(X)) = \text{sn}(J_l(X) + \Theta_l(X)) = \text{sn}(J_l(X))$ при любом $\Theta_l(X) \in \{0, 1\}$. ■

Непосредственное применение теоремы 9 к числу (38) для получения в соответствии с теоремой 8 минимальной ИХМК $\Theta_l(X)$ по (37) к цели не приводит из-за имеющейся неопределенности: $J_l(Z_l(X)) = -1$. В рамках предлагаемой методологии критическая ситуация устраняется с помощью процедуры сужения ИМФ ЦЧ. Эта процедура состоит в преобразовании ИМФ числа $Z_l(X)$ относительно модулей m_1, m_2, \dots, m_l :

$$Z_l(X) = \sum_{i=1}^{l-1} M_{i,l-1} \chi_{i,l-1} + M_{l-1} \hat{I}_l(X) - M_l \quad (41)$$

к его ИМФ относительно модулей m_1, m_2, \dots, m_{l-1} , т. е. к виду

$$Z_l(X) = \sum_{i=1}^{l-2} M_{i,l-2} \chi_{i,l-2} + M_{l-2} \hat{I}_{l-1}(X) + M_{l-1} J_{l-1}(Z_l(X)) \quad (l \in \{3, 4, \dots, k\}). \quad (42)$$

Осуществляемое преобразование базируется на соотношениях (34). Как следует из (35) и (38), ИМФ (41) с помощью (34) приводится к выражению

$$Z_l(X) = \sum_{i=1}^{l-2} M_{i,l-2} \chi_{i,l-2} + M_{l-2} \left(\sum_{i=1}^{l-1} R_{i,l-1}(\chi_i) + m_{l-1} \hat{I}_l(X) - m_{l-1} m_l \right). \quad (43)$$

В соответствии с (6) из (43) для ИИ $I_{l-1}(Z_l(X))$ ЦЧ $Z_l(X)$ относительно модулей m_1, m_2, \dots, m_{l-1} вытекает формула $I_{l-1}(Z_l(X)) = \sum_{i=1}^{l-1} R_{i,l-1}(\chi_i) + m_{l-1} \hat{I}_l(X) - m_{l-1} m_l$. Отсюда согласно определению (18) для интервально-индексных характеристик, входящих в состав ИМФ (42), получаем выражения

$$\hat{I}_{l-1}(Z_l(X)) = |I_{l-1}(Z_l(X))|_{m_{l-1}} = \left| \sum_{i=1}^{l-1} R_{i,l-1}(\chi_i) \right|_{m_{l-1}}; \quad (44)$$

$$J_{l-1}(Z_l(X)) = \lfloor I_{l-1}(Z_l(X)) / m_{l-1} \rfloor = \hat{\rho}_{l-1}(X) + \hat{I}_l(X) - m_l, \quad (45)$$

где $\hat{\rho}_{l-1}(X) = \left[m_{l-1}^{-1} \sum_{i=1}^{l-1} R_{i,l-1}(\chi_i) \right]$; вычеты $R_{i,l-1}(\chi_i)$ вычисляются по формулам типа (14) с заменой l на $l-1$.

Если в результате сужения ИМФ (41), осуществляемого по расчетным соотношениям (44), (45), для всех $i=i_l+1, i_l+2, \dots, l$, где i_l – натуральное число, такое, что $2 < i_l \leq l$, выполняется $J_{i-1}(Z_i(X)) = -1$, а $J_{i_l-1}(Z_{i_l}(X)) \neq -1$ (существование для каждого $l = \overline{3, k}$ указанного i_l гарантируется теоремой 9), то числа $Z_{i_l}(X), Z_{i_l+1}(X), \dots, Z_l(X)$ совпадают, причем согласно теоремам 8 и 9 $\text{sn}(Z_{i_l}(X)) = \text{sn}(J_{i_l-1}(Z_{i_l}(X)))$, а $\Theta_j(X) = 1 - \text{sn}(Z_{i_l}(X))$ ($j=i_l, i_l+1, \dots, l$).

Согласно теоремам 2 и 3 $\rho_{l-1}(X) = \hat{\rho}_{l-1}(X) + \Theta_{l-1}(X) \leq l-2$. Следовательно, при $m_l \geq l-2$ верна оценка $\hat{\rho}_{l-1}(X) \leq l-2$. Поэтому ввиду

$$0 \leq \hat{\rho}_{l-1}(X) + \hat{I}_l(X) \leq l-2 + m_l - 1 \leq 2m_l - 1 \quad (46)$$

детектирование знака главного ИИ (45) равносильно формированию признака

$$\omega_l = \left[(\hat{\rho}_{l-1}(X) + \hat{I}_l(X)) / m_l \right] \quad (47)$$

переполнения при выполнении операции $\hat{x}_l = | \hat{\rho}_{l-1}(X) + \hat{I}_l(X) |_{m_l}$ сумматором по модулю m_l . При этом на неопределенную ситуацию в процедуре сужения ИМФ ЦЧ $Z_l(X)$ указывает единичное значение булевой величины

$$\delta_l = \begin{cases} 0, & \text{если } \hat{x}_l \neq m_l - 1; \\ 1, & \text{если } \hat{x}_l = m_l - 1. \end{cases} \quad (48)$$

Из (46) следует, что в случае, когда $m_l = m_{l-2}$, величина (48) может принимать значение $\delta_l = 1$ не только при $J_{l-1}(Z_l(X)) = -1$ ($\omega_l = 0$), но и при $J_{l-1}(Z_l(X)) = m_l - 1$ ($\omega_l = 1$). Видно, что вторая из указанных ситуаций не возникает, если m_l ограничить снизу порогом $l-1$ (см. (45)–(48)).

На основании изложенных базовых теоретических положений арифметики неизбыточных МСС с диапазонами неотрицательных ЦЧ синтезирован алгоритм расчета ИХМК, который допускает как параллельную, так и последовательную реализации.

Приведем числовые примеры, демонстрирующие реализационные свойства различных форм представления ЦЧ (см. (5), (6), (8)). При этом в качестве базовой будем использовать МСС с четырьмя основаниями ($k=4$): $m_1=2, m_2=3, m_3=5, m_4=7$. Найдем необходимые системные константы:

$$M_4 = 2 \cdot 3 \cdot 5 \cdot 7 = 210;$$

$$M_{1,4} = M_4 / m_1 = 105, M_{2,4} = M_4 / m_2 = 70, M_{3,4} = M_4 / m_3 = 42, M_{4,4} = M_4 / m_4 = M_3 = 30;$$

$$M_{1,3} = M_3 / m_1 = 15, M_{2,3} = M_3 / m_2 = 10, M_{3,3} = M_3 / m_3 = M_2 = 6;$$

$$M_{1,2} = M_2 / m_1 = 3, M_{2,2} = M_2 / m_2 = M_1 = m_1 = 2;$$

$$\left| M_{1,4}^{-1} \right|_{m_1} = \left| 105^{-1} \right|_2 = 1, \left| M_{2,4}^{-1} \right|_{m_2} = \left| 70^{-1} \right|_3 = 1, \left| M_{3,4}^{-1} \right|_{m_3} = \left| 42^{-1} \right|_5 = 3, \left| M_{4,4}^{-1} \right|_{m_4} = \left| M_3^{-1} \right|_{m_4} = \left| 30^{-1} \right|_7 = 4;$$

$$\left| M_{1,3}^{-1} \right|_{m_1} = \left| 15^{-1} \right|_2 = 1, \left| M_{2,3}^{-1} \right|_{m_2} = \left| 10^{-1} \right|_3 = 1, \left| M_{3,3}^{-1} \right|_{m_3} = \left| M_2^{-1} \right|_{m_3} = \left| 6^{-1} \right|_5 = 1;$$

$$\left| M_{1,2}^{-1} \right|_{m_1} = \left| 3^{-1} \right|_2 = 1, \left| M_{2,2}^{-1} \right|_{m_2} = \left| M_1^{-1} \right|_{m_2} = \left| 2^{-1} \right|_3 = 2;$$

$$\begin{aligned} | -m_1^{-1} |_{m_4} &= \left| -\frac{1}{2} \right|_7 = 3, \quad | -m_2^{-1} |_{m_4} = \left| -\frac{1}{3} \right|_7 = 2, \quad | -m_3^{-1} |_{m_4} = \left| -\frac{1}{5} \right|_7 = 4; \\ | -m_1^{-1} |_{m_3} &= \left| -\frac{1}{2} \right|_5 = 2, \quad | -m_2^{-1} |_{m_3} = \left| -\frac{1}{3} \right|_5 = 3; \\ | -m_1^{-1} |_{m_2} &= \left| -\frac{1}{2} \right|_3 = 1, \quad | m_1^{-1} |_{m_2} = \left| \frac{1}{2} \right|_3 = 2. \end{aligned}$$

Пример 1. Сформируем интегрально-характеристические таблицы, которые необходимы для вычисления ранга, ИИ и цифр полиадического кода числа.

Используя полученные константы из (14), находим

$$\begin{aligned} \text{TIC1}_4 &= \{ R_{1,4}(\chi_1) \mid \chi_1 \in \mathbf{Z}_{m_1} \} = \{ | -m_1^{-1} | M_{1,3}^{-1} \chi_1 |_{m_4} \mid \chi_1 \in \mathbf{Z}_{m_1} \} = \{ 3 | \chi_1 |_2 | \chi_1 \in \{0, 1\} \} = \{0, 3\}; \\ \text{TIC2}_4 &= \{ 2 | \chi_2 |_3 | \chi_2 \in \mathbf{Z}_3 \} = \{0, 2, 4\}; \quad \text{TIC3}_4 = \{ 4 | \chi_3 |_5 | \chi_3 \in \mathbf{Z}_5 \} = \{0, 4, 1, 5, 2\}; \\ \text{TIC4}_4 &= \{ M_3^{-1} \chi_4 |_{m_4} \mid \chi_4 \in \mathbf{Z}_{m_4} \} = \{ 4 \chi_4 | \chi_4 \in \mathbf{Z}_7 \} = \{0, 4, 1, 5, 2, 6, 3\}; \\ \text{TIC1}_3 &= \{ 2 | \chi_1 |_2 | \chi_1 \in \{0, 1\} \} = \{0, 2\}; \quad \text{TIC2}_3 = \{ 3 | 2 \chi_2 |_3 | \chi_2 \in \mathbf{Z}_3 \} = \{0, 1, 3\}; \\ \text{TIC3}_3 &= \{ M_2^{-1} \chi_3 |_{m_3} \mid \chi_3 \in \mathbf{Z}_{m_3} \} = \{ 1 \cdot \chi_3 | \chi_3 \in \mathbf{Z}_5 \} = \{0, 1, 2, 3, 4\}; \\ \text{TIC1}_2 &= \{ 1 | \chi_1 |_2 | \chi_1 \in \{0, 1\} \} = \{0, 1\}; \quad \text{TIC2}_2 = \{ 2 \chi_2 | \chi_2 \in \mathbf{Z}_3 \} = \{0, 2, 1\}. \end{aligned}$$

Пример 2. Получим в заданной МСС ранговую форму (5) для числа $X = (1, 0, 1, 6)$. Найдем сначала ранг $\rho_4(X)$ ЦЧ X . Используя полученные в примере 1 таблицы, из (13), (21), (29) имеем

$$\begin{aligned} \hat{I}_4(X) &= \left| \sum_{i=1}^4 \text{TIC}i_4[\chi_i] \right|_{m_4} = |3+0+4+3|_7 = 3; \\ \hat{\rho}_4(X) &= \left\lfloor m_4^{-1} \sum_{i=1}^4 \text{TIC}i_4[\chi_i] \right\rfloor = \lfloor (3+0+4+3)/7 \rfloor = 1; \\ \hat{\rho}_3(X) &= \left\lfloor m_3^{-1} \sum_{i=1}^3 \text{TIC}i_3[\chi_i] \right\rfloor = \lfloor (2+0+1)/5 \rfloor = 0. \end{aligned}$$

Так как ЦЧ $\hat{\rho}_3(X) + \hat{I}_4(X) - m_4 = 0+3-7=-4$ отрицательно, то согласно теоремам 8 и 9 минимальное ИХМК $\Theta_4(X)=0$. Следовательно, по теореме 3 ранг $\rho_4(X) = \hat{\rho}_3(X) + \Theta_4(X) = 1$.

Таким образом, ранговая форма (5) ЦЧ X имеет вид $X = 105 | 1 \cdot \chi_1 |_2 + 70 | 1 \cdot \chi_2 |_3 + 42 | 3 \chi_3 |_5 + 30 | 4 \chi_4 |_7 - 210 \rho_4(X) = 105 + 0 + 126 + 90 - 210 = 111$.

Пример 3. Для числа $X = (1, 0, 1, 6)$ получим ИМФ (6), по теореме 5 ИИ $I_4(X) = \hat{I}_4(X) - m_4 \Theta_4(X) = 3 - 7 \cdot 0 = 3$. С учетом этого согласно (6) $X = 15 | 1 \cdot \chi_1 |_2 + 10 | 1 \cdot \chi_2 |_3 + 6 | 1 \cdot \chi_3 |_5 + 30 I_4(X) = 15 + 0 + 6 + 90 = 111$.

Пример 4. Получим полиадическую форму (8) ЦЧ $X = (1, 0, 1, 6)$. Согласно теореме 7 $x_1 = \chi_1 = 1$, $x_2 = \hat{I}_2(X) = | \text{TIC1}_2[\chi_1] + \text{TIC2}_2[\chi_2] |_{m_2} = | 1 + 0 |_3 = 1$, $x_3 = \hat{x}_3 = | \hat{\rho}_2(X) + \hat{I}_3(X) |_{m_3} = | \lfloor (1+0)/3 \rfloor + | 2 + 0 + 1 |_5 |_3 = 3$. Так как ЦЧ $\hat{\rho}_2(X) + \hat{I}_3(X) - m_3 = 0+3-5 = -2 < 0$, то $\Theta_3(X) = 0$ (см. теоремы 8 и 9). Согласно примеру 2 $\hat{x}_4 = | \hat{\rho}_3(X) + \hat{I}_4(X) |_{m_4} = | 0 + 3 |_7 = 3$. Следовательно, $x_4 = | \hat{x}_4 + \Theta_3(X) |_{m_4} = | 3 + 0 |_7 = 3$.

Таким образом, в рассматриваемом случае (8) имеет вид $X = \sum_{i=1}^4 M_{i-1} \cdot x_i = 1 + 2 \cdot 1 + 6 \cdot 3 + 30 \cdot 3 = 111$.

Приведенные примеры показывают, что среди рассмотренных форм ЦЧ ИМФ занимает приоритетное положение. В первую очередь это обусловлено модульностью расчетного соотношения для оценочного значения $\hat{I}_k(X)$ ИИ $I_k(X)$, что упрощает его вычисление, а следовательно, и синтезируемые на базе ИМФ немодульные процедуры. В наибольшей мере данное свойство интервально-индексных характеристик и ИМФ проявляется при оперировании в диапазонах больших чисел.

Заключение

Исследования по проблематике разработки и оптимизации технологий расчета ИХМК показали следующие результаты:

1. Все известные ИХМК: ранг, ИИ, коэффициенты полиадической формы, интервальный номер ЦЧ – имеют одинаковую вычислительную структуру. Она описывается расчетными соотношениями, в рамках которых исходные значения ИХМК представляются в виде суммы по основанию МСС приближенного (оценочного) значения характеристики и минимальной ИХМК, умноженной на целочисленную константу. Благодаря отмеченному обстоятельству все ИХМК формируются по единому алгоритму.

2. В построенной интегрально-характеристической базе модулярной арифметики ключевая роль принадлежит интервально-индексным характеристикам. Это обусловлено тем, что по ИИ и главному ИИ с помощью тривиальных выражений вычисляются оценочные значения всех других ИХМК. Что касается ранга и ранговой формы ЦЧ, то они используются в качестве вспомогательного средства в процессе математической формализации аппарата ИМФ, а также при выводе определяющего выражения для минимальных ИХМК.

3. Центральное место в разработанной интегрально-характеристической базе МСС занимает процедура сужения ИМФ. Именно она позволяет вычислять все ИХМК, включая минимальные, в рамках общего универсального алгоритма. С точки зрения компьютерной реализации важнейшим свойством процедуры сужения ИМФ является ее параллелизм, порождаемый независимостью друг от друга осуществляемых вычислений по разным модулям базиса МСС.

Разработка выполнена в рамках задания 1.5.06 ГПНИ «Информатика и космос» (2011–2015 гг.).

Список литературы

1. Жуков-Емельянов, О.Д. Информационные технологии на основе модулярной алгебры / О.Д. Жуков-Емельянов. – М. : КРАСАНД, 2010. – 248 с.

2. Параллельная компьютерная алгебра // Всерос. науч. конф. с элементами научной школы для молодежи : сб. науч. тр. Ставрополь, 11–15 окт., 2010 г. – Ставрополь : Издательско-информационный центр «Фабула», 2010. – 364 с.

3. Теоретические основы минимально избыточных квадратичных модулярных систем счисления / А.Ф. Чернявский [и др.] // Доклады НАН Беларуси. – 1998. – Т. 42, № 1. – С. 5–12.

4. Минимально избыточные полиномиально-скалярные модулярные системы счисления / А.А. Коляда [и др.] // Весці НАН Беларусі. Сер. фіз.-мат. навук. – 1998. – № 3. – С. 103–107.

5. Калмыков, И.А. Теоретические основы вычислений в полиномиальной системе классов вычетов, ориентированных на построение отказоустойчивых систем : автореф. дис. ... д-ра техн. наук : 05.13.17, 05.13.15 / И.А. Калмыков. – Ставрополь, 2006. – 32 с.

6. Амербаев, В.М. Теоретические основы машинной арифметики / В.М. Амербаев. – Алма-Ата : Наука, 1976. – 324 с.

7. Cox-Rower architecture for fast parallel Montgomery multiplication / S. Kawamura [et al.] // Eurocrypt 2000, LNCS. – Berlin, 2000. – Vol. 1807. – P. 523–538.
8. Нейрокомпьютеры в остаточных классах. Кн. 11 / Н.И. Червяков [и др.]. – М. : Радиотехника, 2003. – 272 с.
9. Виноградов, И.М. Основы теории чисел / И.М. Виноградов. – М. : Наука, 1972. – 168 с.
10. Коляда, А.А. Модулярные структуры конвейерной обработки цифровой информации / А.А. Коляда, И.Т. Пак. – Минск : Университетское, 1992. – 256 с.
11. Червяков, Н.И. Нейронная сеть для преобразования полиадического кода в код системы остаточных классов / Н.И. Червяков, Д.В. Сивоплясов, Д.В. Горденко // Нейрокомпьютеры: разработка, применение. – 2003. – № 10–11. – С. 10–12.
12. Преобразователь из модулярного кода в обобщенную полиадическую систему счисления для отказоустойчивых систем управления / И.А. Калмыков [и др.] // Успехи современного естествознания. – 2009. – № 4. – С. 41–43.

Поступила 25.09.2012

*Институт прикладных физических проблем
им. Севченко БГУ,
Минск, Курчатова, 7
e-mail: shabinskaya@rambler.ru*

A.A. Kolyada, A.F. Chernyavsky

INTEGRATED CHARACTERISTIC BASE OF MODULAR NUMBER SYSTEMS

The problem of development of an integrated characteristic base of a modular number system defined on the ranges of non-negative integers is studied. A mechanism of interval-modular forms of integer numbers is applied to solve this problem. Interval index characteristics – the interval index and the main interval index - play a key role in this mechanism. Priority of these characteristics is explained by their essential advantages over known integrated characteristics of a modular code when optimizing algorithms of non-modular operations.

ЛОГИЧЕСКОЕ ПРОЕКТИРОВАНИЕ

УДК 519.7

Ю.В. Поттосин, С.Н. Кардаш

ПОВЫШЕНИЕ БЫСТРОДЕЙСТВИЯ КОМБИНАЦИОННОЙ СХЕМЫ ПУТЕМ КОНВЕЙЕРИЗАЦИИ

Рассматривается вопрос повышения быстродействия устройства без памяти, преобразующего последовательность дискретных сигналов. Поставлена задача разбиения заданной многоуровневой комбинационной схемы на заданное число каскадов, на выходах которых должны быть поставлены регистры для обеспечения конвейерной обработки поступающих сигналов. Для решения этой задачи используется модель, основанная на представлении комбинационной схемы в виде ориентированного графа. В процессе решения задачи определяется частота поступления сигналов на вход схемы, и эта частота по возможности должна быть максимальной.

Введение

Повышению производительности систем обработки информации всегда уделялось большое внимание. Одним из способов повышения производительности является применение структуры конвейерного типа, которая имеет еще название «трубопровод» (пер. с англ. pipeline) [1]. Подобную структуру образуют несколько независимых процессоров, соединенных между собой так, что информация на выходе одного процессора является входной информацией для другого процессора. Процессоры образуют информационный конвейер. Выходной процессор выдает результаты через короткие интервалы времени, хотя действительное время прохождения потока информации через конвейер может быть довольно большим.

Принцип конвейеризации эффективно используется, когда характер обработки информации представляется как последовательность операций, каждая из которых состоит из последовательности этапов [2, 3]. Для того чтобы начать выполнение последующей операции, не надо ждать окончания всего процесса выполнения предыдущей операции. Достаточно, чтобы у предыдущей операции был закончен только первый этап. Если конвейер имеет r последовательных этапов, $(i + k)$ -я операция может проходить $(s - k)$ -й этап, где $1 \leq s$, $s - k \leq r$, в то время когда i -я операция проходит s -й этап.

При построении систем цифровой обработки сигналов в режиме реального времени широкое распространение получил систолический принцип организации вычислений, основанный на сочетании идей реализации параллелизма и конвейеризации вычислений с учетом тенденций развития технологии СБИС [4, 5]. Сама технология ориентирована на уменьшение стоимости изготовления одного элемента на кристалле СБИС. Разрабатывается один унифицированный элемент СБИС, и из элементов одного типа строится цепочка, матрица или другая структура. Элементы работают параллельно, выполняя базовую операцию. После ее выполнения осуществляется синхронная передача выходных данных от одного элемента к соседним по всем локальным связям.

Информация в систолическом процессоре распространяется по конвейеру подобно тому, как пульсирует кровь при сокращении систолы сердца. Аналогия между выполнением базовой операции в элементе структуры и сокращением сердца, а также между пульсациями потока исходных данных, промежуточных и окончательных результатов и кровотоков определила название данного класса процессоров как систолических.

В данной работе предпринята попытка найти способ повышения быстродействия путем конвейеризации многоуровневой комбинационной схемы, построенной на основе СБИС.

1. Постановка задачи

В многоуровневой схеме устройства задержка складывается из задержек элементов самой длинной цепочки. Пусть на вход комбинационной схемы поступает последовательность p наборов

двоичных сигналов. Если T – время задержки схемы, то период смены сигналов не может быть меньше T . Время реакции устройства на данную последовательность в этом случае будет не меньше pT . Разобьем схему на k каскадов (C_1, C_2, \dots, C_k), и если τ_C – время задержки самого медленно действующего каскада, то $T \leq k\tau_C$. На выходы каждого каскада поставим элементы задержки (триггеры D), пропускающие сигналы с выходов каскада по сигналу синхронизации. Этот же сигнал синхронизации определяет период смены сигналов на входе устройства, который должен быть не меньше суммы двух задержек: задержки τ_C и задержки τ_D элемента D ($\tau_{\text{clock}} \geq \tau_C + \tau_D$). Теперь время реакции устройства на упомянутую последовательность длины p равно $(k + p)\tau_{\text{clock}}$.

Пример разделения схемы на три каскада показан на рис. 1, где границы каскадов обозначены штриховыми линиями. В схеме использованы модули, содержащиеся в библиотеке КМОП-элементов проектирования заказных СБИС [6].

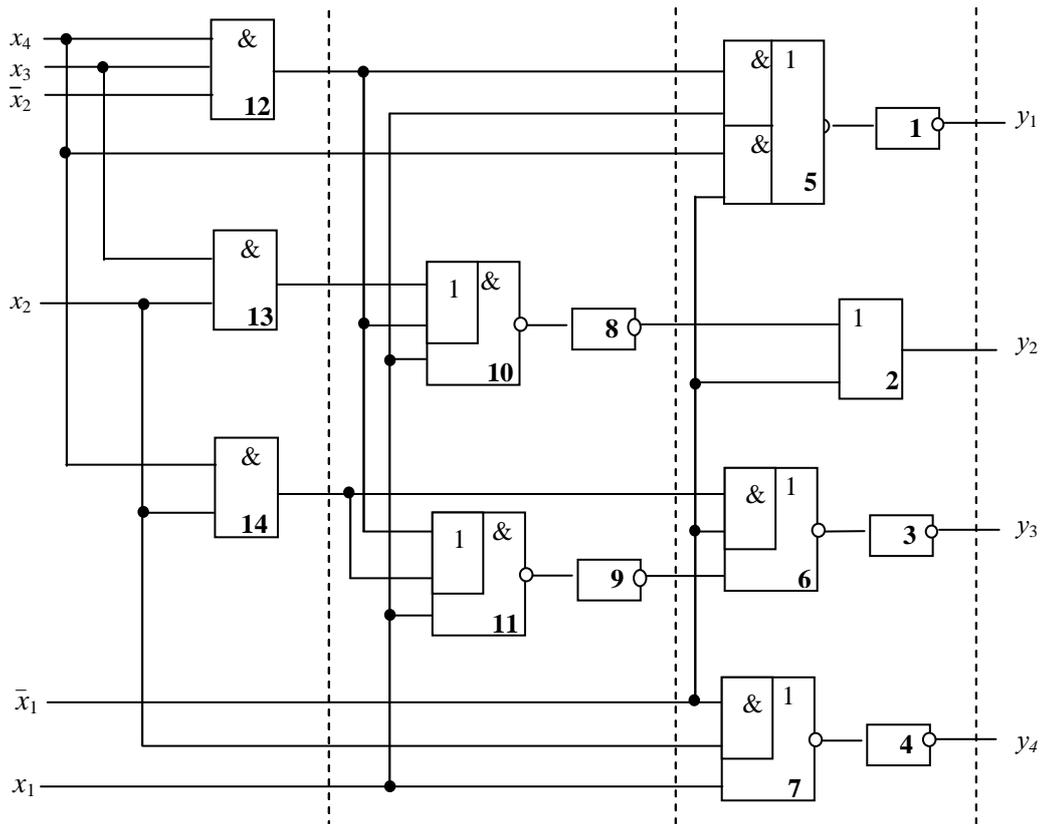


Рис. 1. Схема с выделенными каскадами

Нижняя граница длины p последовательности наборов двоичных сигналов на входе устройства, при которой имеет место ускорение обработки сигналов, определяется неравенством

$$(k + p)\tau_{\text{clock}} < pT, \tag{1}$$

где T – время задержки исходной схемы. Учитывая установленную выше нижнюю границу периода следования сигналов синхронизации $\tau_{\text{clock}} \geq \tau_C + \tau_D$, получим

$$p > \frac{k\tau_{\text{clock}}}{T - \tau_{\text{clock}}} \geq \frac{k(\tau_C + \tau_D)}{T - \tau_C - \tau_D}. \tag{2}$$

Заданную комбинационную схему требуется разбить на требуемое число k каскадов, чтобы обеспечить максимальное быстродействие при описанном конвейерном режиме.

2. Модель комбинационной схемы

Для представления комбинационной схемы удобно использовать в качестве модели ориентированный граф (орграф) $G = (V, A)$ с множеством вершин V и множеством дуг A (рис. 2). Вершины орграфа представляют логические элементы и входные полюсы схемы (индексы у символов вершин совпадают с номерами соответствующих элементов схемы на рис. 1), а дуги показывают направления сигналов от выходов одних элементов к входам других.

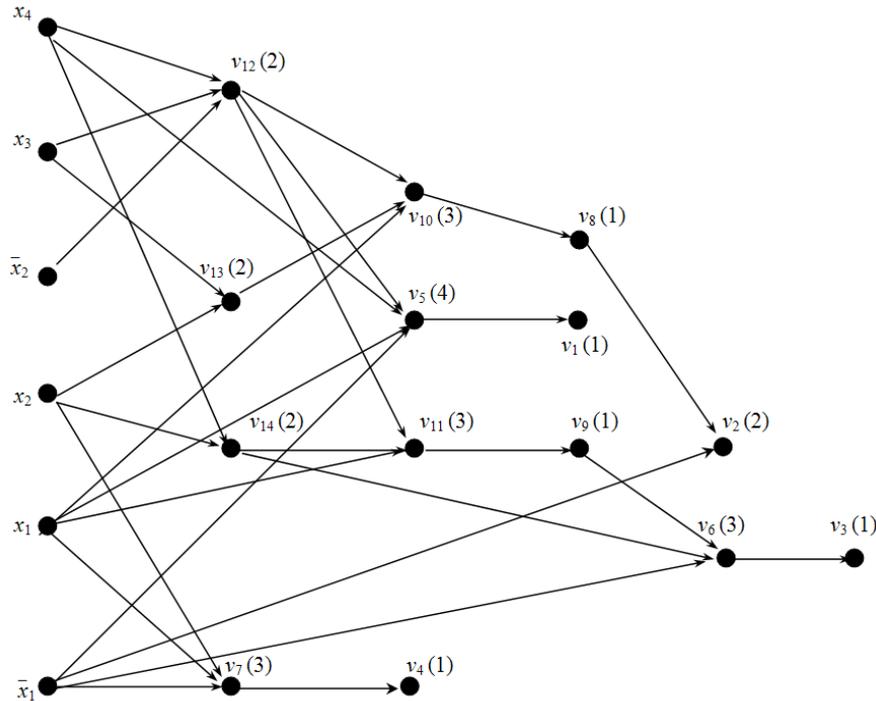


Рис. 2. Орграф G

Орграф G не содержит контуров. Каждой его вершине $v \in V$ приписан вес $\tau(v)$, представляющий задержку соответствующего элемента. Весами могут быть целые числа, пропорциональные временам задержки элементов. Вершины, соответствующие входам схемы, имеют вес, равный нулю. Пусть вершины орграфа на рис. 2 указаны в скобках. При этом считаем, что чем сложнее булева функция, реализуемая логическим элементом, тем большую задержку имеет данный элемент.

3. Разбиение схемы на слои

Сформируем последовательность слоев L_1, L_2, \dots, L_m , представляющую собой упорядоченное разбиение множества вершин V орграфа G с таким свойством, что если вершина v принадлежит полуокрестности исхода $N^+(u)$ вершины u , то эти вершины находятся в разных слоях и слой, содержащий вершину u , предшествует в этой последовательности слою с вершиной v (не обязательно непосредственно). Если длины путей от входов схемы к ее выходам различны, то данное разбиение не является единственным. Следует выбрать такой вариант разбиения на слои, чтобы сумма весов всех слоев была по возможности минимальной. Под весом слоя понимаем максимум весов вершин, принадлежащих этому слою.

Можно выделить два типа вершин орграфа G . К одному типу отнесем вершины, которые лежат на самых длинных путях в орграфе G . Они строго распределяются по слоям и не могут менять свое положение. Назовем их *неподвижными*. Положение в слоях других вершин, которые назовем *подвижными*, можно менять в определенных пределах, скажем, от слоя L_l до слоя L_r ($l < r$). Эти пределы устанавливаются довольно легко. Достаточно выполнить приведенный

ниже алгоритм 1 для заданного орграфа G и орграфа G^c , полученного из G изменением направлений всех дуг. В алгоритме 1 приняты следующие обозначения: $N^-(v)$ – полуокрестность захода, $N^+(v)$ – полуокрестность исхода вершины v , L_i – i -й слой, m – число слоев.

Алгоритм 1

- 1) $L_1 := \{v / N^-(v) = \emptyset\}$, $i := 1$;
- 2) $i := i + 1$, $L_i := \bigcup_{v \in L_{i-1}} N^+(v)$. Если $L_i \neq \emptyset$, перейти к 2,
иначе $j := i$, $m := i := i - 1$;
- 3) $i := j := j - 1$. Если $j = 1$, перейти к 5,
иначе
- 4) $i := i - 1$. Если $i = 1$, перейти к 3,
иначе $L_i := L_i \setminus L_j$, перейти к 4;
- 5) конец.

После выполнения алгоритма 1 над оргграфом G^c необходимо порядок в полученной последовательности слоев поменять на обратный. Неподвижные вершины окажутся в одних и тех же слоях. Для подвижной вершины «левым» слоем L_l является тот, который получен при выполнении алгоритма 1 над оргграфом G , «правым» слоем L_r – тот, который получен при выполнении алгоритма 1 над оргграфом G^c .

На рис. 2 и 3 показаны распределения вершин по слоям в результате применения алгоритма 1 к оргграфу G (см. рис. 2) и к оргграфу G^c (рис. 3). На обоих рисунках слои представлены вертикальными рядами вершин. Для орграфа G получены слои $L_1 = \{x_1, \bar{x}_1, x_2, \bar{x}_2, x_3, x_4\}$, $L_2 = \{v_7, v_{12}, v_{13}, v_{14}\}$, $L_3 = \{v_4, v_5, v_{10}, v_{11}\}$, $L_4 = \{v_1, v_8, v_9\}$, $L_5 = \{v_2, v_6\}$, $L_6 = \{v_3\}$, а для орграфа G^c после изменения порядка слоев – $L_1 = \{x_2, \bar{x}_2, x_3, x_4\}$, $L_2 = \{x_1, v_{12}, v_{14}\}$, $L_3 = \{v_{11}, v_{13}\}$, $L_4 = \{\bar{x}_1, v_9, v_{10}\}$, $L_5 = \{v_5, v_6, v_7, v_8\}$, $L_6 = \{v_1, v_2, v_3, v_4\}$. Неподвижными являются вершины $\bar{x}_2, x_2, x_3, x_4, v_3, v_6, v_9, v_{11}, v_{12}$ и v_{14} . С неподвижными вершинами образовалось шесть слоев $L_1 = \{x_2, \bar{x}_2, x_3, x_4\}$, $L_2 = \{v_{12}, v_{14}\}$, $L_3 = \{v_{11}\}$, $L_4 = \{v_9\}$, $L_5 = \{v_6\}$ и $L_6 = \{v_3\}$ соответственно с весами 0, 2, 3, 1, 3 и 1.

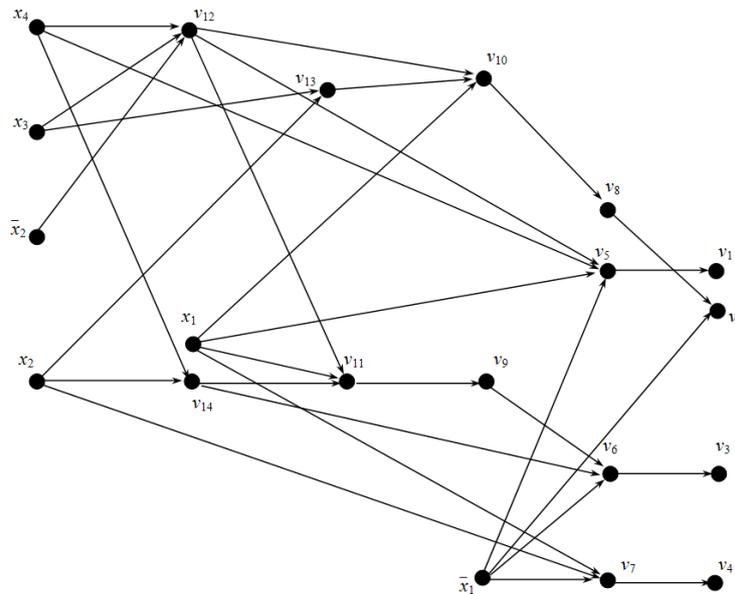


Рис. 3. Распределение по слоям вершин орграфа G , полученное применением алгоритма 1 к оргграфу G^c

Остальные вершины являются подвижными, и для каждой из них, как показано выше, существует непустое множество слоев, где вершина может находиться. Вершина x_1 может находиться в слоях L_1 и L_2 , вершина \bar{x}_1 – в слоях L_1, L_2, L_3 и L_4 , вершина v_1 – в сло-

ях L_4 , L_5 и L_6 , вершина v_2 – в слоях L_5 и L_6 , вершина v_4 – в слоях L_3 , L_4 , L_5 и L_6 , вершина v_5 – в слоях L_3 , L_4 и L_5 , вершина v_7 – в слоях L_2 , L_3 , L_4 и L_5 , вершина v_8 – в слоях L_4 и L_5 , вершина v_{10} – в слоях L_3 и L_4 , вершина v_{13} – в слоях L_2 и L_3 . Однако положение одних вершин зависит от положения других. В частности, вершины, связанные дугой, не могут находиться в одном слое.

Для окончательного распределения вершин по слоям так, чтобы сумма весов слоев была по возможности минимальной, предлагается следующий способ. Удалив из орграфа G неподвижные вершины вместе с инцидентными им ребрами, получим оргграф H , в каждой компоненте которого выделим вершину с максимальным весом. Эту вершину поместим в один из допустимых для нее слоев с максимальным весом. Границы положения вершин при этом изменятся, и некоторые вершины из подвижных перейдут в неподвижные. Дальнейшее распределение по слоям можно вести для каждой компоненты орграфа H описанным выше способом.

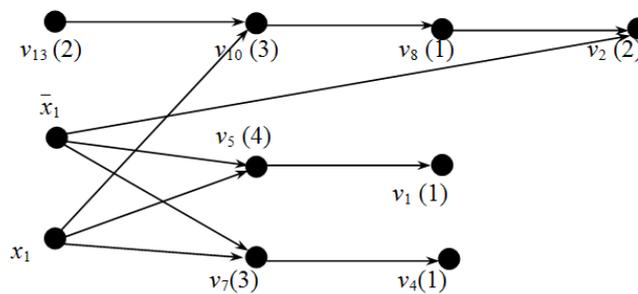


Рис. 4. Оргграф H , полученный из орграфа G

Для рассматриваемого примера оргграф H , состоящий из одной компоненты, показан на рис. 4. Вершины x_1 и \bar{x}_1 с нулевым весом могут быть помещены в любой слой. Вершину v_5 с максимальным весом согласно указанному способу поместим в слой L_3 . Окончательно получим $L_1 = \{x_1, \bar{x}_1, x_2, \bar{x}_2, x_3, x_4\}$, $L_2 = \{v_{12}, v_{13}, v_{14}\}$, $L_3 = \{v_5, v_7, v_{10}, v_{11}\}$, $L_4 = \{v_1, v_4, v_8, v_9\}$, $L_5 = \{v_2, v_6\}$ и $L_6 = \{v_3\}$.

4. Выравнивание путей в орграфе

Приведем все пути в графе к единой длине путем добавления новых вершин с нулевым весом. В результате исходный оргграф $G = (V, A)$ преобразуется в оргграф $G' = (V', A')$. В оргграф G добавляются новые вершины так, чтобы выполнялось $N^+(v) \subseteq L_{i+1}$ для любых $v \in L_i$ и $i = 1, 2, \dots, m-1$ при сохранении достижимости вершин. Под сохранением достижимости понимаем то, что во всяком пути из вершины $v \in V$ к вершине $u \in V$ в орграфе G' встречаются те же вершины из V в том же порядке, что и в соответствующем пути в орграфе G . Новые вершины снабжаются нулевым весом. Для этого используется алгоритм 2. Преобразованный таким образом оргграф изображен на рис. 5, где добавленные вершины обозначены светлыми кружками.

Алгоритм 2

- 1) $i := 0, j := 0; U := \emptyset;$
- 2) $i := i + 1$. Если $i = m$, перейти к 4,
иначе $L := L_i;$
- 3) если $L = \emptyset$, перейти к 2,
иначе выбрать $v \in L, L := L \setminus \{v\}, A := N^+(v) \setminus L_{i+1}$. Если $A = \emptyset$, перейти к 3,
иначе $j := j + 1, U := U \cup \{u_j\}, N^+(v) := (N^+(v) \cap L_{i+1}) \cup \{u_j\}, N^+(u_j) := A,$
 $L_{i+1} := L_{i+1} \cup \{u_j\}$, перейти к 3;
- 4) конец.

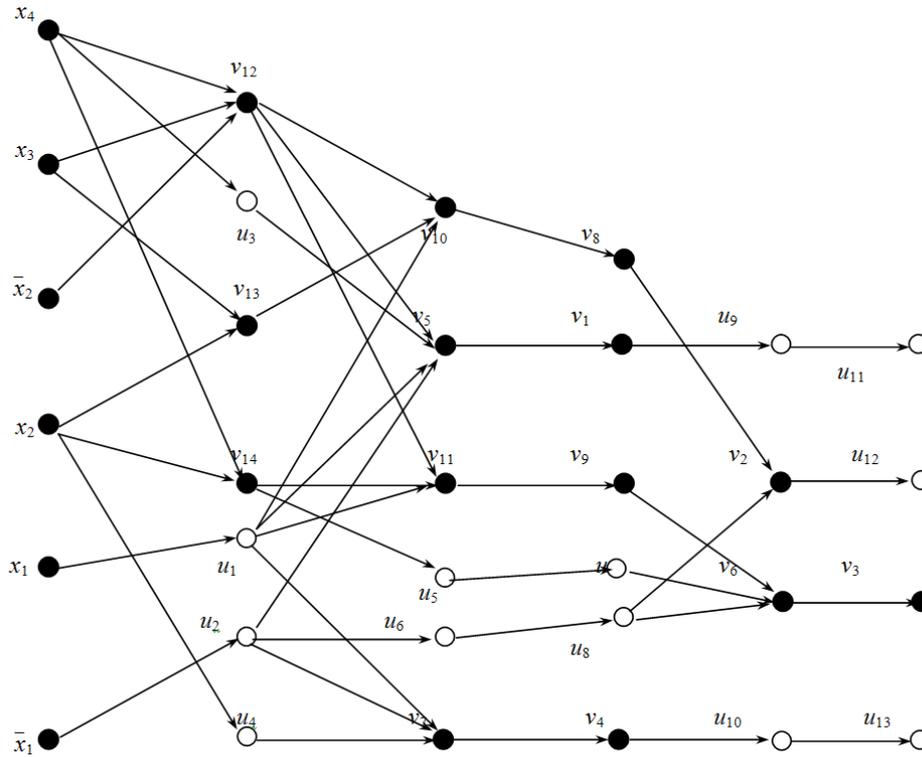


Рис. 5. Орграф G' с выравненными длинами путей

5. Разбиение схемы на каскады

Каждому из слоев соответствует множество значений веса, приписанных вершинам, которые принадлежат данному слою. Максимальное значение веса в этом слое представляет собой задержку прохождения сигнала в слое. Заданную комбинационную схему надо разбить на заданное число каскадов с минимизацией задержки в самом медленно действующем каскаде. Каждый каскад представляет собой упорядоченное множество слоев. Возникает следующая задача.

Дана последовательность положительных чисел (a_1, a_2, \dots, a_n) . Назовем отрезком последовательности ее часть вида $(a_p, a_{p+1}, \dots, a_q)$, где $1 \leq p < q \leq n$. Требуется разбить заданную последовательность на заданное число k отрезков B_1, B_2, \dots, B_k , где $B_i = (a_{n_{i-1}+1}, \dots, a_{n_i})$, $i = 1, 2, \dots, k$, $n_0 = 0, n_1, n_2, \dots, n_k = n$, которым соответствуют S_1, S_2, \dots, S_k , где $S_i = \sum_{a_j \in B_i} a_j$, так, чтобы величина $\max\{S_1, S_2, \dots, S_k\}$ была минимальной.

Элементы B_i данной последовательности соответствуют слоям в заданной схеме, а каждый из них равен максимуму веса в соответствующем слое.

Предлагается следующий метод получения решения данной задачи, близкого к оптимальному. Сначала определяется нижняя граница максимальной задержки в каскаде как

$$b = (\sum_{j=1}^n a_j) / n.$$

Очередной i -й отрезок формируется путем накопления суммы $S_i = \sum_{a_j \in B_i} a_j$, кото-

рая сравнивается с b . При $S_i > b$ после добавления очередного a_j к S_i , если $b - (S_i - a_j) > S_i - b$ (недобор до b больше перебора), то крайним правым элементом в отрезке B_i устанавливается a_j , в противном случае — a_{j-1} . Та же процедура повторяется для оставшейся неразделенной части последовательности.

Ниже представлен алгоритм 3, получающий решение для случая, когда $(\sum_{j=1}^k S_j) / k > a_i$,

для любого $i = 1, 2, \dots, n$. Решение этой задачи представим в виде последовательности n_1, n_2, \dots, n_k , где n_i – номер крайнего правого элемента в i -м отрезке. В данном алгоритме τ – максимальная задержка в каскаде, S – среднее значение S_i , B – текущее значение S_i .

Алгоритм 3

- 1) $i := 0, R := 0, \tau := 0, l := k, n_k := n$;
- 2) если $i = n$, перейти к 3,
иначе $i := i + 1, R := R + a_i$, перейти к 2;
- 3) $S := R / l, i := 0, B := 0, j := 1$;
- 4) $i := i + 1$. Если $i = n$, то перейти к 6,
иначе $C := B, B := B + a_i$. Если $S > B$, перейти к 4,
иначе $D := S - C, E := B - S$. Если $D < E$, то $j := j + 1, n_j := i - 1, B := a_i$, перейти к 5,
иначе $j := j + 1, n_j := i, C := B, B := 0$;
- 5) $l := l - 1, R := R - C, S := R / l$. Если $\tau < C$, то $\tau := C$, перейти к 4,
иначе перейти к 4;
- 6) конец.

В рассматриваемом примере последовательность чисел (0, 2, 4, 1, 3, 1) представляет задержки в слоях. Допустим, эту последовательность надо разбить на три отрезка ($k = 3, n = 6$). Результатом выполнения алгоритма 3 является последовательность $(n_1, n_2, n_3) = (2, 3, 6)$, а максимальная задержка в каскаде – 5.

Триггеры в схеме должны стоять после второго, третьего и шестого слоев, на выходе каждого из крайних правых элементов каждого каскада. Соответствующим образом преобразованный орграф показан на рис. 6, где вершины, соответствующие триггерам, помечены символами d_i . Вершины, представляющие фиктивные элементы с нулевой задержкой, должны быть удалены из орграфа.

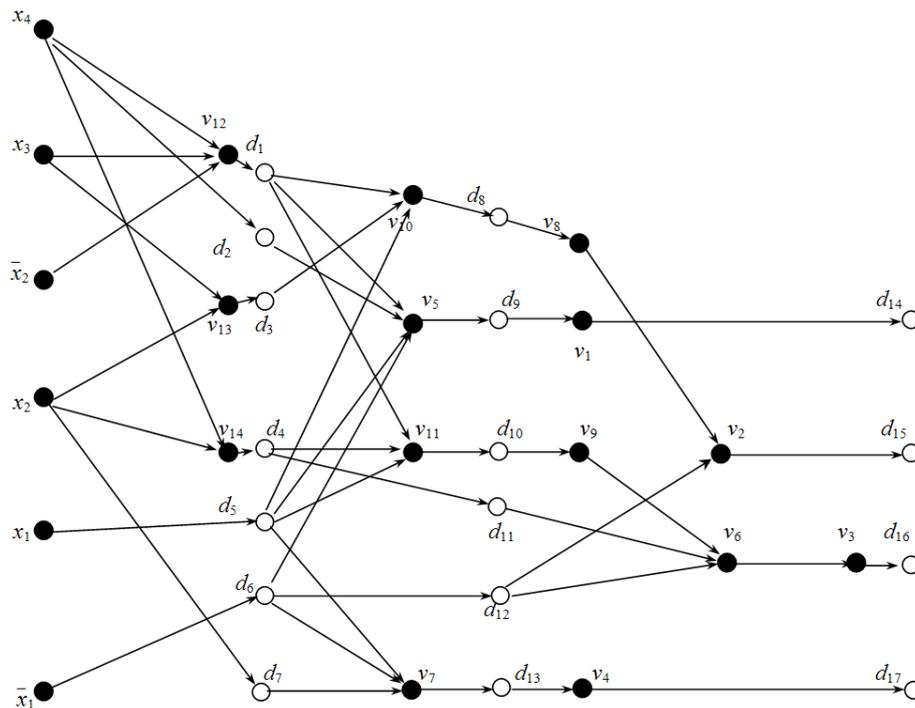


Рис. 6. Орграф с вершинами, соответствующими триггерам

Результат преобразования схемы, представленной на рис. 1, в схему, работающую в режиме конвейеризации, изображен на рис. 7, где вместо инверторов используются инверсные выходы триггеров.

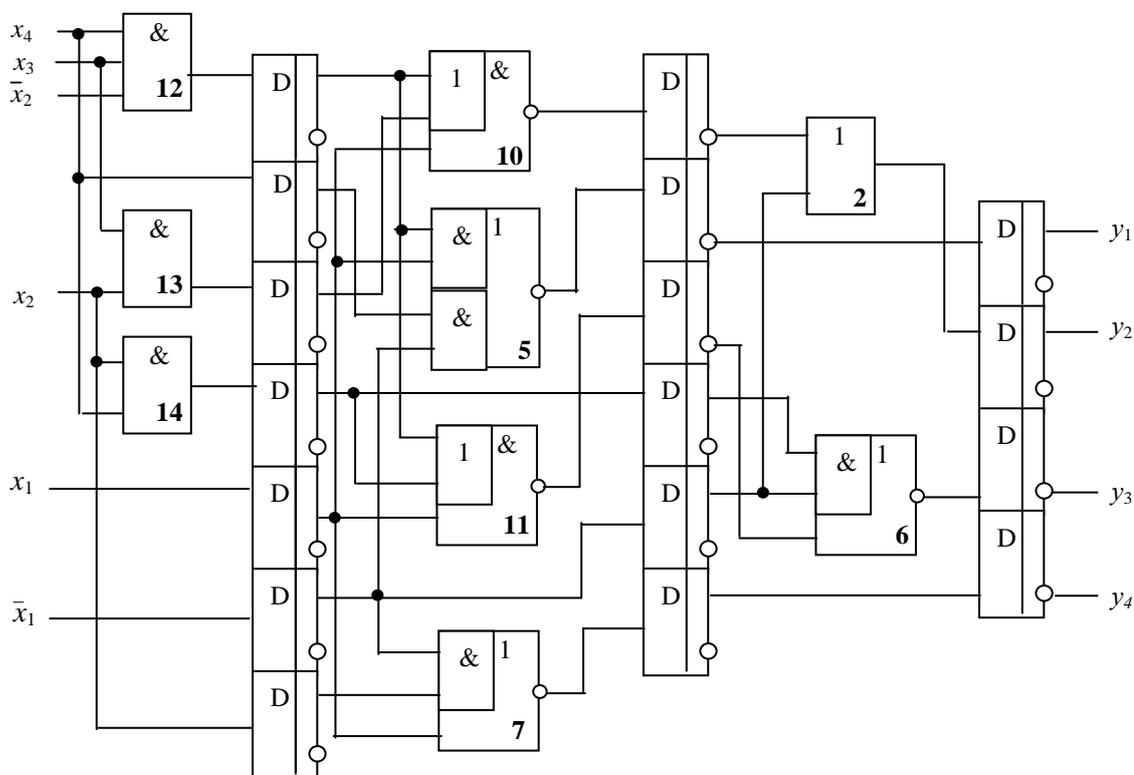


Рис. 7. Схема, работающая в конвейерном режиме

Допустим, что задержка триггера составляет четыре условные единицы, используемые при оценке задержки в каскадах схемы из приведенного примера. Задержка самого медленного каскада равна пяти. Следовательно, длительность такта синхронизации должна быть не менее девяти. Согласно формуле (2) ускорение обработки сигналов может быть достигнуто при длине входной последовательности не менее 14. Полная реакция исходной схемы на такую последовательность будет через 154 единицы времени, тогда как на такую же последовательность конвейеризованная схема выдаст реакцию через 153 единицы времени. Из формулы (1) видно, что разница между этими величинами будет увеличиваться с увеличением длины входной последовательности.

Заключение

Предлагаемый подход к повышению быстродействия комбинационной схемы рассчитан на применение его к готовым схемам и решает данную задачу на схемотехническом уровне. Возможно, на функциональном уровне можно добиться более значительного эффекта. Однако в этом случае необходимо заново проектировать схему исходя из ее функционального описания.

Список литературы

1. Каган, Б.М. Цифровые вычислительные машины и системы / Б.М. Каган, М.М. Каневский. – М. : Энергия, 1973. – 680 с.
2. Воеводин, В.В. Математические модели и методы в параллельных процессах / В.В. Воеводин. – М. : Наука. Гл. ред. физ.-мат. лит., 1986. – 296 с.
3. Капитонова, Ю.В. Математическая теория проектирования вычислительных систем / Ю.В. Капитонова, А.А. Летичевский. – М. : Наука. Гл. ред. физ.-мат. лит., 1988. – 296 с.

4. Кухарев, Г.А. Системные процессоры для обработки сигналов / Г.А. Кухарев, А.Ю. Тропченко, В.П. Шмерко. – Минск : Беларусь, 1988. – 127 с.
5. Кухарев, Г.А. Алгоритмы и системные процессоры для обработки многозначных данных / Г.А. Кухарев, В.П. Шмерко, Е.Н. Зайцева. – Минск : Навука і тэхніка, 1990. – 296 с.
6. Лукошко, Г. КМОП-базовые матричные кристаллы серии К1574 / Г. Лукошко, Е. Коннов // Радиолучитель. – 1997. – № 9. – С. 39–40.

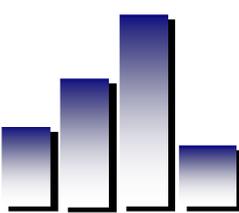
Поступила 28.11.12

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

Yu.V. Pottosin, S.N. Kardash

INCREASING COMBINATIONAL CIRCUIT PERFORMANCE VIA PIPELINING

The question of increasing performance of a device with no memory, which develops a sequence of discrete signals, is considered. A problem is set to divide a given multilevel combinational circuit into a given number of cascades with registers providing pipeline-wise development of incoming signals. To solve this problem we use a model based on representation of combinational circuit as a directed graph. In the process of solving this problem, the frequency of incoming signals is established. This frequency must be as high as possible.



**Computer
Data
Analysis &
Modeling** 10th International Conference

2013 Theoretical & Applied
Minsk, Belarus Stochastics

September, 10–14, 2013

**FIRST ANNOUNCEMENT
& CALL FOR PAPERS**

In the International Year of Statistics, this Conference will provide a forum for researchers and a discussion of the latest results in theory and software of data analysis, statistical modeling and computer simulation, stressing on Theoretical and Applied Stochastics, and define ways for further developments in this field. Young researchers will have the possibility to present their results and to get in contact with experienced scientists. The conference is organized by the Belarusian State University, Research Institute for Applied Problems of Mathematics and Informatics, the Institute of Mathematics of the National Academy of Sciences of Belarus, the United Institute of Informatics Problems of the National Academy of Sciences of Belarus, the Belarusian Statistical Association, and the Vienna University of Technology, Institute of Statistics and Probability Theory.

The main topics of the Conference:

1. Robust and Nonparametric Data Analysis;
2. Multivariate Analysis and Design of Experiments;
3. Statistical Analysis of Time Series and Random Fields;
4. Probabilistic and Statistical Analysis of Discrete Data;
5. Asymptotic Methods in Stochastics;
6. Statistical Methods of Signal and Image Processing;
7. Econometric Modeling and Financial Mathematics;
8. Survey Analysis and Official Statistics;
9. Computer Simulation of Stochastic Systems;
10. Computer Intensive Methods, Algorithms and Statistical Software;
11. Computer Data Analysis in Applications

Conference site: Belarusian State University, Minsk, Belarus

Conference language: English.

DEADLINES

February	28, 2013	– Preliminary registration on the Conference Web-site
March	15, 2013	– Submission of papers for reviewing
April	10, 2013	– Notification on acceptance of papers
April	20, 2013	– Submission of camera ready manuscripts
April	20, 2013	– Final registration
September	10, 2013	– Conference desk opens

Proceedings will be published before the Conference.

Submission of papers. To prepare your paper the LaTeX word processor and the macros downloaded from <http://www.cdam.bsu.by> only must be used! Also on that web site the guidelines are available. The paper, up to 4 full pages of A4 macros format, including text, figures, and references, should be sent by e-mail to Dr. Alexey Kharin: KharinAY@bsu.by. After the reviewing process the Proceedings will be published by the BSU Publishing House, and will be distributed at the Registration Desk.

Special issues: Extended versions of the selected papers will be published in Special Issues of the “Austrian Journal of Statistics”, the “International Journal Informatica” and the “Russian Journal of Applied Econometrics”.

Conference web-site: <http://www.cdam.bsu.by>