

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ УКРАИНЫ

ХАРЬКОВСКИЙ НАЦИОНАЛЬНЫЙ УНИВЕРСИТЕТ
РАДИОЭЛЕКТРОНИКИ

ISSN 1563-0064

РАДИОЭЛЕКТРОНИКА И ИНФОРМАТИКА

Научно-технический журнал

Основан в 1997 г.

№ 1(72), январь – март 2016

Выходит 4 раза в год

© Харьковский национальный
университет радиоэлектроники, 2016

Свидетельство о государственной регистрации КВ № 12097-968 ПР 14.12.2006

СОДЕРЖАНИЕ

РАДИОТЕХНИКА

ВЛАСЕНКО Л.А., РУТКАС А.Г. ОБ ОПТИМАЛЬНОМ УПРАВЛЕНИИ ПЕРЕХОДНЫМИ ПРОЦЕССАМИ В ЭЛЕКТРИЧЕСКИХ ЦЕПЯХ.....	3
---	---

ЭЛЕКТРОНИКА

ОКСАНИЧ И.Г., ПРИТЧИН А.С., КАСАТКИН А.В. МАТЕМАТИЧЕСКАЯ МОДЕЛЬ ДЛЯ ВИРТУАЛЬНОГО МОНИТОРИНГА РАДИАЛЬНОЙ НЕРАВНОМЕРНОСТИ РАСПРЕДЕЛЕНИЯ ПРИМЕСИ В РАСПЛАВЕ И СЛИТКЕ КРЕМНИЯ НА ПРИМЕРЕ ФОСФОРА.....	8
---	---

ТЕЛЕКОММУНИКАЦИИ

ЮДИН А.К., БАРАННИК В.В., ФРОЛОВ О.В. ТЕХНОЛОГИЯ ПОВЫШЕНИЯ БЕЗОПАСНОСТИ ИНФОРМАЦИОННЫХ РЕСУРСОВ НА ОСНОВЕ ИСПОЛЬЗОВАНИЯ ФУНКЦИОНАЛЬНОГО ПРЕОБРАЗОВАНИЯ ПРИ КОСВЕННОМ СТЕГАНОГРАФИЧЕСКОМ ВСТРАИВАНИИ.....	14
--	----

СИСТЕМЫ И ПРОЦЕССЫ УПРАВЛЕНИЯ

ПИЧУГИНА О.С., ЯКОВЛЕВ В.С. МЕТОД ШТРАФНЫХ ФУНКЦИЙ ДЛЯ РЕШЕНИЯ ЗАДАЧ ОПТИМИЗАЦИИ НА ПОЛИЭДРАЛЬНО- СФЕРИЧЕСКИХ КОМБИНАТОРНЫХ МНОЖЕСТВАХ.....	18
--	----

НОВОЖИЛОВА М.В., ЛИТВИНЕНКО Є.М. ДИНАМІЧНА МОДЕЛЬ ОПТИМІЗАЦІЇ ТРАНСПОРТНИХ ВИТРАТ З УРАХУВАННЯМ НЕВИЗНАЧЕНОСТІ ТЕРМІНІВ ДОСТАВКИ.....	27
--	----

ПРОХОРОВ В. П., ПРОХОРОВ А. В. ПРИНЯТИЕ РЕШЕНИЙ В СИСТЕМЕ КОНТРОЛЯ И АНАЛИЗА ПАРАМЕТРОВ ДВИЖЕНИЯ АВТОТРАНСПОРТА.....	32
---	----

КОМПЬЮТЕРНЫЕ НАУКИ

ГРАБОВСЬКА Н.Р., РУСИН Б.П., ІВАНЮК В.Г. ПОХИБКА ТРИВИМІРНОЇ РЕКОНСТРУКЦІЇ ПОВЕРХНІ ЗА ТРИАДОЮ ЇЇ ЗОБРАЖЕНЬ.....	39
---	----

КОМПЬЮТЕРНАЯ ИНЖЕНЕРИЯ И ТЕХНИЧЕСКАЯ ДИАГНОСТИКА

ТАМЕР ВАНІ АМЕР, ЧУМАЧЕНКО С.В., ЕМЕЛЬЯНОВ И.В. КУБИТНАЯ ФОРМА ОПИСАНИЯ ВЫЧИСЛИТЕЛЬНЫХ СТРУКТУР.....	47
---	----

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

НАРОЖНЫЙ В.В., НАРОЖНАЯ Е.В. ИССЛЕДОВАНИЕ КАЧЕСТВА СБОРКИ ПРИЛОЖЕНИЙ ПОД ОПЕРАЦИОННУЮ СИСТЕМУ ANDROID РАЗЛИЧНЫМИ КОМПИЛЯТОРАМИ.....	53
--	----

ЛЕВЫКИН В.М., ЧАЛАЯ О.В. ФОРМАЛИЗАЦИЯ НЕЯВНОГО ПРОЦЕДУРНОГО ЗНАНИЯ	57
--	----

БАРАННИК В.В., ПОДЛЕСНЫЙ С.А., ШУЛЬГИН С.С. МЕТОДОЛОГИЯ ОЦЕНКИ ВЛИЯНИЯ КИБЕРАТАК НА БЕЗОПАСНОСТЬ ВИДЕОИНФОРМАЦИОННОГО РЕСУРСА В ТЕЛЕКОММУНИКАЦИОННЫХ СИСТЕМАХ.....	61
--	----

РЕФЕРАТЫ.....	65
---------------	----



**ОБОПТИМАЛЬНОМУПРАВЛЕНИИ
ПЕРЕХОДНЫМИПРОЦЕССАМИВ
ЭЛЕКТРИЧЕСКИХЦЕПЯХ**

ВЛАСЕНКО Л.А., РУТКАС А.Г.

Изучается задача оптимального управления переходными процессами в электрических цепях с сосредоточенными линейными элементами. Качество управления характеризуется квадратичный функционал энергии, зависящий от параметров цепи. Переходные процессы моделируются линейными дифференциально-алгебраическими уравнениями, для которых получены условия существования и единственности оптимального управления. Результаты применяются для расчета оптимального управляющего напряжения в цепи.

Ключевые слова: оптимальное управление, переходной процесс, электрическая цепь, дифференциально-алгебраическое уравнение.

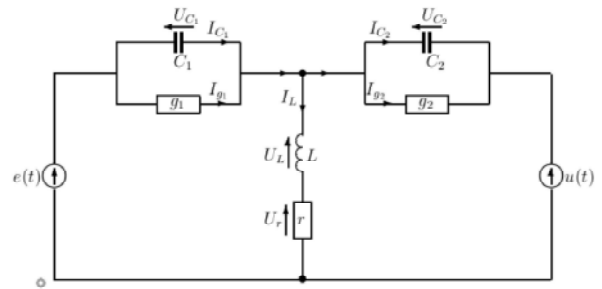
Key words: optimal control, transient process, electrical circuit, differential-algebraic equation.

1. Введение

Рассмотрим задачу оптимального управления в электрических цепях, математические модели которых описываются с помощью дифференциально-алгебраических уравнений. Для этого изучим линейно-квадратичную задачу оптимального управления в системе, динамика которой описывается дифференциально-алгебраическим уравнением. Такие системы управления также называют дескрипторными. Существует обширная литература по задаче оптимального управления для дескрипторных систем (см., например, [1-4]). Настоящая работа продолжает исследования, начатые в [4]. В отличие от работы [4] здесь не налагаются никакие ограничения на резольвенту характеристического пучка матриц, отвечающего уравнению. Ограничения также сняты в задаче импульсного управления в [5].

2. Постановка задачи для электрической цепи

Для демонстрации предлагаемого метода оптимального управления рассмотрим электрический четырехполюсник, изображенный на рисунке. Четырехполюсник имеет два источника напряжения $e(t)$ и $u(t)$; $e(t)$ – заданная функция от времени, $u(t)$ – управляемый источник напряжения. Параллельно с емкостями C_1, C_2 включены проводимости g_1, g_2 . На внутренней ветви расположены индуктивность L и сопротивление r .



Электрический четырехполюсник

Токи $I_{C_1}, I_{C_2}, I_{g_1}, I_{g_2}, I_L$ и напряжения $U_{C_1}, U_{C_2}, U_r, U_L$ удовлетворяют законам Кирхгофа

$$\begin{aligned} I_{C_1} + I_{g_1} &= I_{C_2} + I_{g_2} + I_L, \\ U_{C_1} + U_L + U_r &= E, \quad U_{C_1} + U_{C_2} + u = E, \end{aligned} \quad (1)$$

и уравнениям

$$\begin{aligned} U_L &= L \frac{d}{dt} I_L, \quad U_r = r I_L, \\ I_{C_j} &= C_j \frac{d}{dt} U_{C_j}, \quad I_{g_j} = g_j U_{C_j}, \quad j=1,2, \end{aligned} \quad (2)$$

где L, r, C_j, g_j – положительные постоянные. Состояние электрической цепи характеризуется вектором

$$x(t) = (x_1(t), x_2(t), x_3(t))^T = (I_L(t), U_{C_1}(t), U_{C_2}(t))^T,$$

состоящим из «энергетических» компонент, которые отвечают инерционным элементам. С помощью (1),(2) получаем систему трех дифференциально-алгебраических уравнений относительно «энергетических» компонент:

$$\begin{aligned} L \frac{d}{dt} I_L + U_{C_1} + r I_L &= e(t), \\ C_1 \frac{d}{dt} U_{C_1} - C_2 \frac{d}{dt} U_{C_2} - I_L + g_1 U_{C_1} - g_2 U_{C_2} &= 0, \\ U_{C_1} + U_{C_2} &= e(t) - u(t). \end{aligned} \quad (3)$$

Перепишем систему уравнений (3) в векторной форме относительно вектора состояний $x(t)$:

$$\frac{d}{dt} [Ax(t)] + Bx(t) = f(t) + Ku(t), \quad (4)$$

где

$$\begin{aligned} A &= \begin{pmatrix} L & 0 & 0 \\ 0 & C_1 & -C_2 \\ 0 & 0 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} r & 1 & 0 \\ -1 & g_1 & -g_2 \\ 0 & 1 & 1 \end{pmatrix}, \\ f(t) &= \begin{pmatrix} e(t) \\ 0 \\ e(t) \end{pmatrix}, \quad K = \begin{pmatrix} 0 \\ 0 \\ -1 \end{pmatrix}. \end{aligned} \quad (5)$$

Задача оптимального управления в системе (3) состоит в выборе входного (управляющего) напряжения $u(t)$, реализующего минимум «энергетического» квадратичного функционала качества этой системы. Вид функционала уточним в разделе 4 данной статьи.

3. Постановка и решение задачи оптимального управления для дифференциально-алгебраического уравнения

В данном разделе относительно уравнения (4) будем предполагать: A, B – вещественные матрицы размеров $n \times n$, K – вещественная матрица размера $n \times m$, $f(t)$ – n -мерная вектор-функция со значениями в \mathbf{R}^n . Управление системой (4) осуществляется с помощью m -мерного вектора управления $u(t)$, которому отвечает n -мерный вектор состояния системы $x(t) = x(t; u)$. Уравнение (4) будем рассматривать на отрезке времени $[0, T]$. Начальное условие зададим в виде

$$Ax(0) = q. \quad (6)$$

Чтобы исследовать задачу оптимального управления в системе (4),(6), опишем решения $x(t) = x(t; u)$ этой системы при различных допустимых управлениях.

Введем обозначения: $L_2(0, T; \mathbf{R}^n)$ – пространство вектор-функций со значениями в \mathbf{R}^n , суммируемых с квадратом нормы на $[0, T]$; $W_2^k(0, T; \mathbf{R}^n)$ – пространство Соболева порядка k вектор-функций из $L_2(0, T; \mathbf{R}^n)$, у которых обобщенные производные до порядка k включительно принадлежат $L_2(0, T; \mathbf{R}^n)$. Функции из $W_2^1(0, T; \mathbf{R}^n)$ будем считать непрерывными на $[0, T]$, изменив их, если необходимо, на множестве нулевой меры. Символ $\langle \cdot, \cdot \rangle$ обозначает скалярное произведение в пространстве \mathbf{R}^n или \mathbf{R}^m , символ E обозначает единичную матрицу соответствующей размерности, а транспонированную матрицу к матрице K обозначаем через K' .

Пусть $f(t) \in L_2(0, T; \mathbf{R}^n)$, $u(t) \in L_2(0, T; \mathbf{R}^m)$. Функцию $x(t) \in L_2(0, T; \mathbf{R}^n)$ назовем решением задачи (4),(6) на отрезке $[0, T]$, если $Ax(t) \in W_2^1(0, T; \mathbf{R}^n)$, функция $x(t)$ почти всюду удовлетворяет уравнению (4) и выполнено начальное условие (6). Управлению $u(t)$ отвечает решение $x(t) = x(t; u)$ системы (4),(6).

На динамику системы существенно влияет пучок матриц $\lambda A + B$. Будем предполагать, что этот пучок матриц является регулярным, т.е. его определитель $\det(\lambda A + B)$ отличен от тождественного нуля. Тогда в некоторой окрестности бесконечно удаленной точки $|\lambda| \geq \vartheta$ нет собственных чисел пучка $\lambda A + B$ (или нулей многочлена $\det(\lambda A + B)$ в комплексной плоскости). Определим матрицы:

$$Q_1 = \frac{1}{2\pi i} \oint_{|\lambda|=\vartheta} A(\lambda A + B)^{-1} d\lambda, \quad Q_2 = E - Q_1,$$

$$G = Q_1 A + Q_2 B, \quad W = -BG^{-1}Q_1, \quad F = Q_2 A G^{-1},$$

где матрица F является нильпотентной с индексом нильпотентности ν ($F^\nu = 0, F^{\nu-1} \neq 0$). Заметим, что матрицы Q_1, Q_2 являются вещественными, хотя контурное интегрирование осуществляется по комплексной переменной. Будем предполагать, что $FK = 0$, и если $\nu > 1$, то $F^k f(t) \in W_2^k(0, T; \mathbf{R}^n)$ для $k = 1, \dots, \nu - 1$. Рассуждая, как при доказательстве теоремы 1 из [6], получаем, что для любого вектора q из начального условия (6) такого, что

$$Q_2 q = \sum_{k=0}^{\nu-1} (-1)^k \frac{d^k}{dt^k} [F^{k+1} f(t)]_{t=0}, \quad (7)$$

существует единственное решение $x(t)$ начальной задачи (4),(6) на отрезке $[0, T]$ и это решение допускает представление

$$x(t) = G^{-1} \left\{ e^{Wt} Q_1 q + \int_0^t e^{W(t-\tau)} Q_1 [f(\tau) + Ku(\tau)] d\tau \right\} + G^{-1} Q_2 \left\{ \sum_{k=0}^{\nu-1} (-1)^k \frac{d^k}{dt^k} [F^k f(t)] + Ku(t) \right\}.$$

Подобный результат для более общего дифференциально-операторного уравнения можно найти в [7].

Задача оптимального управления заключается в оптимизации управления $u(t) \in L_2(0, T; \mathbf{R}^m)$, реализующего минимум функционала качества

$$\min_{u \in L_2} J(u),$$

$$J(u) = \int_0^T \left[\langle N_1 x(t), x(t) \rangle_{\mathbf{R}^n} + \langle N_2 u(t), u(t) \rangle_{\mathbf{R}^m} \right] dt \quad (8)$$

на решениях $x(t) = x(t; u)$ задачи (4),(6), где N_1, N_2 – квадратные неотрицательно определенные матрицы соответственно размеров n, m , причем $N_2 \geq \delta I$, $\delta > 0$.

Управление $u_*(t)$, на котором достигается минимум функционала (8), будем называть оптимальным управлением, а соответствующее решение $x_*(t) = x(t; u_*)$ – оптимальным решением. Будем придерживаться схемы метода, предложенной в [4]. Существование и единственность оптимального управления устанавливает

Теорема 1. Пусть пучок матриц $\lambda A + B$ является регулярным; $f(t) \in L_2(0, T; \mathbf{R}^n)$; если индекс нильпотентности ν матрицы F больше 1, то $F^k f(t) \in W_2^k(0, T; \mathbf{R}^n)$ для $k = 1, \dots, \nu - 1$; матрица K

при управлении в уравнении (4) такая, что $FK = 0$. Тогда для любого начального вектора q в условиях (6), удовлетворяющем соотношению (7), существует единственное оптимальное управление $u_*(t)$, минимизирующее функционал качества (8).

Доказательство. Доказательство теоремы 1 осуществляется по схеме доказательства теоремы 2 из [4]. Единственное решение $x(t; u)$, задачи (4),(6) и функционал $J(u)$ (8) представимы в виде

$$x(t; u) = L(u)(t) + w(t),$$

$$J(u) = \langle N_1(Lu + w), Lu + w \rangle + \langle N_2u, u \rangle,$$

где L – ограниченный линейный оператор из $L_2(0, T; \mathbf{R}^m)$ в $L_2(0, T; \mathbf{R}^n)$ вида

$$(Lu)(t) = G^{-1} \int_0^t e^{W(t-\tau)} Q_1 Ku(\tau) d\tau + G^{-1} Q_2 Ku(t),$$

$$w(t) = G^{-1} \left\{ e^{Wt} Q_1 q + \int_0^t e^{W(t-\tau)} Q_1 f(\tau) d\tau \right\} + G^{-1} Q_2 \sum_{k=0}^{v-1} (-1)^k \frac{d^k}{dt^k} [F^k f(t)].$$

К оператору L определим сопряженный оператор L^* из $L_2(0, T; \mathbf{R}^n)$ в $L_2(0, T; \mathbf{R}^m)$:

$$(L^*v)(t) = K'Q_1' \int_t^T e^{W(\tau-t)} (G')^{-1} v(\tau) d\tau + K'Q_2' (G')^{-1} v(t). \quad (9)$$

Оператор $M = N_2 + L^*N_1L$ является ограниченным линейным оператором в $L_2(0, T; \mathbf{R}^m)$, который самосопряжен и имеет ограниченный обратный, норма которого оценивается как $\|M^{-1}\| \leq 1/\delta$. Функция

$$u_*(t) = -(M^{-1}L^*N_1w)(t) \quad (10)$$

является единственным оптимальным управлением задачи (4),(6),(8), так как непосредственные вычисления показывают, что $J(u) - J(u_*) \geq \delta \|u - u_*\|^2$ для всех $u \in L_2(0, T; \mathbf{R}^m)$. На этом доказательство теоремы завершается.

Согласно подходу, принятому в [8], введем в рассмотрение сопряженное состояние $p(t)$ и сопряженную систему, решением которой является сопряженное состояние. Для этого заметим, что соотношение

$$N_2u + L^*N_1x(u) = 0 \quad (11)$$

выполняется тогда и только тогда, когда $u = u_*$ является оптимальным управлением и $x(u_*) = x_*$ является оптимальным решением (ср. с утверждением 1 [4]).

Используя вид сопряженного оператора L^* (9), равенство (11) переписываем в форме

$$N_2u + K'Q_1' \int_t^T e^{W(\tau-t)} (G')^{-1} N_1x(\tau) d\tau + K'Q_2' (G')^{-1} N_1x(t) = 0.$$

Теперь можно ввести сопряженное состояние

$$p(t) = Q_1' \int_t^T e^{W(\tau-t)} (G')^{-1} N_1x(\tau) d\tau + Q_2' (G')^{-1} N_1x(t). \quad (12)$$

Тогда оптимальное управление (10) вычисляется через сопряженное состояние (12) по формуле:

$$u_*(t) = -N_2^{-1}K'p(t). \quad (13)$$

Нетрудно проверить, что функция

$$Q_1' p(t) = \int_t^T e^{W(\tau-t)} Q_1' (G')^{-1} N_1x(\tau) d\tau$$

является единственным решением задачи

$$\frac{d}{dt}[Q_1' p(t)] + W'[Q_1' p(t)] + Q_1' (G')^{-1} N_1x(t) = 0, \text{ п.в. } t \in [0, T], \quad Q_1' p(T) = 0,$$

таким, что $Q_1' p(t) \in W_2^1(0, T; \mathbf{R}^n)$. Отсюда следует, что функция $p(t)$ (12) является единственным решением задачи

$$\frac{d}{dt}[A'Q_1' p(t)] = B'p(t) - N_1x(t), \quad (14)$$

п.в. $t \in [0, T]$, $Q_1' p(T) = 0$,

таким, что $A'Q_1' p(t) \in W_2^1(0, T; \mathbf{R}^n)$. Согласно соотношению (13) оптимальное решение удовлетворяет следующим соотношениям:

$$\frac{d}{dt}[Ax(t)] + Bx(t) + KN_2^{-1}K'p(t) = f(t), \quad (15)$$

п.в. $t \in [0, T]$, $Ax(0) = q$.

Таким образом, мы получили следующий результат.

Теорема 2. Пусть пучок матриц $\lambda A + B$ является регулярным; $f(t) \in L_2(0, T; \mathbf{R}^n)$; если индекс нильпотентности v матрицы F больше 1, то $F^k f(t) \in W_2^k(0, T; \mathbf{R}^n)$ для $k = 1, \dots, v-1$; матрица K при управлении в уравнении (4) такая, что $FK = 0$;

начальный вектор q в условии (6) удовлетворяет соотношению (7). Тогда задача (14),(15) имеет единственное решение $p(t) \in L_2(0, T; \mathbf{R}^n)$, $x(t) = x_*(t) \in L_2(0, T; \mathbf{R}^n)$. Оптимальное управление $u_*(t) \in L_2(0, T; \mathbf{R}^m)$ строится по формуле (13).

Заметим также, что подход, в котором анализируются задачи оптимального управления с точки зрения функционального анализа, изложен в [9].

4. Приложение результатов раздела 3 к исследованию задачи оптимального управления входным напряжением в электрических цепях

Вернемся к системе (3), описанной в разделе 2. Начальные условия (6) предполагают задание в начальный момент времени следующих значений:

$$LI_L(0) = q_1, \quad C_1 U_{C_1}(0) - C_2 U_{C_2}(0) = q_2. \quad (16)$$

Задача оптимального управления в системе (3),(16) состоит в выборе входного напряжения $u(t) \in L_2(0, T)$, реализующего минимум функционала энергии инерционных элементов и управления:

$$J(u) = \int_0^T [LI_L^2(t) + C_1 U_{C_1}^2(t) + C_2 U_{C_2}^2(t) + u^2(t)] dt.$$

Эта задача допускает представление в виде (4),(6),(8) с параметрами (5) и

$$q = \begin{pmatrix} q_1 \\ q_2 \\ 0 \end{pmatrix}, \quad N_1 = \begin{pmatrix} L & 0 & 0 \\ 0 & C_1 & 0 \\ 0 & 0 & C_2 \end{pmatrix}, \quad N_2 = 1.$$

Пучок матриц $\lambda A + B$ является регулярным. Для несобственных чисел пучка, т.е.

$$\lambda \neq (-a \pm \sqrt{a^2 - 4b})/2, \\ a = L \frac{g_1 + g_2}{C_1 + C_2} + \frac{r}{L}, \quad b = \frac{r(g_1 + g_2) + 1}{L(C_1 + C_2)},$$

существует обратная матрица

$$(\lambda A + B)^{-1} = \\ = d^{-1}(\lambda) \begin{pmatrix} d_2(\lambda) + d_3(\lambda) & -1 & -d_3(\lambda) \\ 1 & d_1(\lambda) & d_1(\lambda)d_3(\lambda) \\ -1 & -d_1(\lambda) & d_1(\lambda)d_2(\lambda) + 1 \end{pmatrix},$$

$$d_1(\lambda) = \lambda L + r, \quad d_{j+1}(\lambda) = \lambda C_j + g_j, \quad j = 1, 2, \\ d(\lambda) = d_1(\lambda)[d_2(\lambda) + d_3(\lambda)] + 1.$$

Находим матрицы

$$Q_1 = \begin{pmatrix} 1 & 0 & \frac{-C_2}{C_1 + C_2} \\ 0 & 1 & \frac{C_1 g_2 - C_2 g_1}{C_1 + C_2} \\ 0 & 0 & 0 \end{pmatrix}, \quad Q_2 = \begin{pmatrix} 0 & 0 & \frac{C_2}{C_1 + C_2} \\ 0 & 0 & \frac{C_2 g_1 - C_1 g_2}{C_1 + C_2} \\ 0 & 0 & 1 \end{pmatrix}.$$

Задача (14) принимает вид

$$\frac{d}{dt} [A' p(t)] = B' p(t) - N_1 x(t), \\ \text{п.в. } t \in [0, T], \quad A' p(T) = 0. \quad (17)$$

В силу теоремы 2 задача (15),(17) имеет единственное решение $x(t) = x_*(t) \in L_2(0, T; \mathbf{R}^3)$, $p(t) \in L_2(0, T; \mathbf{R}^3)$, оптимальное управление строится по формуле $u_*(t) = p_3(t)$. Сделав замену переменных

$$p_3(t) = e(t) - x_2(t) - x_3(t), \quad y(t) = C_1 x_2(t) - C_2 x_3(t), \\ x_2(t) = \frac{y}{C_1 + C_2} + C_2 \alpha p_1(t) + C_2 \beta p_2(t) + \alpha e(t), \\ x_3(t) = \frac{-y}{C_1 + C_2} + C_1 \alpha p_1(t) + C_1 \beta p_2(t) + \alpha \frac{C_1}{C_2} e(t), \\ \alpha = \sigma C_2, \quad \beta = \sigma(C_2 g_1 - C_1 g_2), \\ \sigma = \frac{1}{(C_1 + C_2)(C_1 + C_2 + C_1 C_2)},$$

задачу (15),(17) можно упростить:

$$\frac{d}{dt} \begin{pmatrix} x_1(t) \\ y(t) \\ p_1(t) \\ p_2(t) \end{pmatrix} = \Lambda \begin{pmatrix} x_1(t) \\ y(t) \\ p_1(t) \\ p_2(t) \end{pmatrix} + \varphi(t), \quad \text{п.в. } t \in [0, T], \quad (18)$$

$$x_1(0) = \frac{q_1}{L}, \quad y(0) = q_2, \quad p_1(T) = p_2(T) = 0,$$

$$\Lambda = \begin{pmatrix} -\frac{r}{L} & -1 & \frac{-C_2 \alpha}{L} & \frac{-C_2 \beta}{L} \\ 1 & \frac{-g_1 - g_2}{C_1 + C_2} & -C_2 \beta & \gamma \\ -1 & 0 & \frac{r}{L} & \frac{-1}{L} \\ 0 & \frac{-1}{C_1 + C_2} & \frac{1}{C_1 + C_2} & \frac{g_1 + g_2}{C_1 + C_2} \end{pmatrix},$$

$$\varphi(t) = \begin{pmatrix} \varphi_1(t) \\ \varphi_2(t) \\ \varphi_3(t) \\ \varphi_4(t) \end{pmatrix} = \begin{pmatrix} \frac{C_1 + C_1 C_2}{L(C_1 + C_2 + C_1 C_2)} e(t) \\ -(C_1 + C_2) \beta e(t) \\ 0 \\ 0 \end{pmatrix}, \\ \gamma = \beta(C_1 g_2 - C_2 g_1).$$

Решение задачи (18) запишем с помощью формулы Коши:

$$\begin{pmatrix} x_1(t) \\ y(t) \\ p_1(t) \\ p_2(t) \end{pmatrix} = \exp(\Lambda t) \begin{pmatrix} q_1 \\ L \\ q_2 \\ p_1(0) \\ p_2(0) \end{pmatrix} + \int_0^t \exp(\Lambda(t-\tau)) \varphi(\tau) d\tau,$$

где неизвестные значения $p_1(0), p_2(0)$ найдем с помощью соотношения

$$\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} q_1 \\ L \\ q_2 \\ p_1(0) \\ p_2(0) \end{pmatrix} + \int_0^T \exp(\Lambda(T-\tau)) \begin{pmatrix} \varphi_1(\tau) \\ \varphi_2(\tau) \\ \varphi_3(\tau) \\ \varphi_4(\tau) \end{pmatrix} d\tau = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

5. Выводы

На примере электрического четырехполюсника, изображенного на рисунке, в работе поставлена задача оптимального управления в системе, описывающей переходные режимы в электрических цепях. Качество управления характеризует квадратичный функционал энергии. Так как математические модели переходных режимов описываются линейными дифференциально-алгебраическими уравнениями, то для этих уравнений изучается задача оптимального управления с квадратичным функционалом качества. Получены новые теоремы существования и единственности оптимального управления, из формулировок и доказательств которых вытекают конструктивные алгоритмы и явные формулы для вычисления оптимального управления и оптимального состояния. Показано, как эти теоремы применяются в модельной электрической

цепи для расчета входного (управляющего) напряжения, минимизирующего функционал энергии.

Литература: 1. *Bender D. J. and Laub A. J.* The linear-quadratic optimal regulator for descriptor systems // IEEE Trans. Automat. Control. 1987. Vol. AC-32, No. 6. P. 672–688. 2. *Kunkel P., Mehrmann V.* The linear quadratic optimal control problem for linear descriptor systems with variable coefficients. // Math. Control Signals Systems. 1997. Vol. 10. P. 247–264. 3. *Kurina G.A., Mörz R.* On linear-quadratic optimal control problems for time-varying descriptor systems // SIAM J. Control Optim. 2004. Vol. 42, No. 6. P. 2062–2077. 4. *Власенко Л.А., Руткас А.Г.* Об оптимальном управлении системами, описываемыми неявными эволюционными уравнениями // Дифференц. уравнения. 2009. Т. 45, №3. С. 416-424. 5. *Власенко Л.А., Руткас А.Г., Самойленко А.М.* Проблема импульсного регулятора для одной динамической системы типа Соболева // Укр. мат. журн. 2008. Т. 60, № 8. С. 1027-1034. 6. *Власенко Л.А.* Импульсные дифференциально-алгебраические уравнения в математических моделях электрических цепей // Радиоэлектроника и информатика. 2004. № 32. С. 27-31. 7. *Власенко Л.А., Руткас А.Г.* О дифференциальной игре в системе, описываемой неявным дифференциально-операторным уравнением // Дифференц. уравнения. 2015. Т. 51, № 6. С. 785-795. 8. *Лионс Ж.-Л.* Оптимальное управление системами, описываемыми уравнениями с частными производными. М.: Мир, 1972. 415 с. 9. *Балакришнан А.В.* Прикладной функциональный анализ. М.: Наука, 1980. 384 с.

Рецензент: д-р техн. наук, проф. Кривуля Г.Ф.

Поступила в редколлегию 14.01.2016

Власенко Лариса Андреевна, д-р техн. наук, профессор кафедры прикладной математики Харьковского национального университета им. В.Н. Каразина. Научные интересы: математическое моделирование, дифференциальные уравнения. Адрес: Украина, 61001, Харьков, ул. Плехановская, 2/5, кв. 29, тел.: (057) 732 28 35.

Руткас Анатолий Георгиевич, д-р физ.-мат. наук, профессор кафедры прикладной математики Харьковского национального университета им. В.Н. Каразина. Научные интересы: математическое моделирование, дифференциальные уравнения. Адрес: Украина, 61001, Харьков, ул. Плехановская, 2/5, кв. 29, тел.: (057) 732 28 35.



МАТЕМАТИЧЕСКАЯ МОДЕЛЬ ДЛЯ ВИРТУАЛЬНОГО МОНИТОРИНГА РАДИАЛЬНОЙ НЕРАВНОМЕРНОСТИ РАСПРЕДЕЛЕНИЯ ПРИМЕСИ В РАСПЛАВЕ И СЛИТКЕ КРЕМНИЯ НА ПРИМЕРЕ ФОСФОРА

ОКСАНИЧ И.Г., ПРИТЧИН А.С., КАСАТКИН А.В.

Усовершенствуется математическая модель для виртуального мониторинга радиальной и осевой неравномерности распределения примеси в расплаве и слитке кремния. В результате апробации модели путем сравнения расчетных и экспериментальных данных получено хорошее согласование между измеренной экспериментальным методом и рассчитанной радиальной концентрацией легирующей примеси в слитках кремния.

Введение

Качество процесса изготовления монокристаллов полупроводников и качество конечного продукта во многом зависит от возможности контроля таких технологических параметров, значения которых не могут быть измерены прямыми методами. Это касается равномерности распределения легирующих примесей в теле слитка.

Кремниевые (Si) подложки являются наиболее широко используемым материалом для изготовления электронных приборов, а слитки, выращенные по методу Чохральского, в основном используются для изготовления этих подложек. Для достижения требуемых электрофизических свойств необходимо в процессе выращивания ввести в слиток кремния электрически активные примеси (лигатуру), такие, например, как фосфор или мышьяк для получения полупроводников n-типа или бор для полупроводников p-типа. Концентрация легирующей примеси непосредственно влияет на такие свойства как пороговое напряжение, удельное сопротивление и напряжение пробоя полупроводниковых приборов.

Вследствие особенностей метода Чохральского при выращивании слитка примесь внедряется в слиток неравномерно по его радиусу (так называемая радиальная неоднородность) и неравномерно по оси слитка (осевая неоднородность). Осевая неоднородность вызывает необходимость обрезки слитка с торцов, что приводит к уменьшению выхода годной продукции, а радиальная неоднородность приводит к браку подложки.

Актуальной является задача определения распределения примеси в процессе выращивания слитка как в осевом, так и в радиальном направлении. Непосредственное измерение этого параметра с применением аппаратных сенсоров не представляется возможным, следовательно, решить эту задачу можно только с использованием информационных технологий виртуального мониторинга. Этого можно достичь разработкой математических моделей для расчета распределения примеси в режиме реального времени.

Распределить примесь вдоль направления роста (оси слитка) можно с помощью уравнения кристаллизации [1]. В то же время для определения распределения легирующей примеси в радиальном направлении используются только экспериментальные методы измерений на выращенных кристаллах [2, 3]. Экспериментальные измерения распределений легирующей примеси в подложках кремния, выращенных по методу Чохральского, показали, что существует большая радиальная неоднородность распределения легирующей примеси [4-6]. Общий недостаток информации о радиальной неоднородности распределения примеси, в сочетании с ограниченной информацией по аналитическим методам расчета этого параметра, позволяет предположить, что требуется дополнительный анализ механизма радиального распределения концентрации примеси и параметров процесса выращивания слитков, которые влияют на этот процесс.

Материал и результаты исследований

Расплав кремния, в котором находится легирующая примесь, можно рассматривать как вязкую жидкость, которая хорошо описывается уравнением Навье-Стокса. Уравнения движения расплава при этом выглядят следующим образом:

$$V_r \frac{\partial V_r}{\partial r} + \frac{V_\phi^2}{r} + V_z \frac{\partial V_r}{\partial z} = -\frac{1}{\rho} \frac{\partial p}{\partial r} + v \left(\frac{\partial^2 V_r}{\partial r^2} + \frac{1}{r} \frac{\partial V_r}{\partial r} - \frac{V_r}{r^2} + \frac{\partial^2 V_r}{\partial z^2} \right), \quad (1)$$

$$V_r \frac{\partial V_z}{\partial r} + V_z \frac{\partial V_z}{\partial z} = -\frac{1}{\rho} \frac{\partial p}{\partial z} + v \left(\frac{\partial^2 V_z}{\partial r^2} + \frac{1}{r} \frac{\partial V_z}{\partial r} + \frac{\partial^2 V_z}{\partial z^2} \right) + g\beta\Delta T, \quad (2)$$

$$V_r \frac{\partial V_\phi}{\partial r} + \frac{V_r V_\phi}{r} + V_z \frac{\partial V_\phi}{\partial z} = v \left(\frac{\partial^2 V_\phi}{\partial r^2} + \frac{1}{r} \frac{\partial V_\phi}{\partial r} - \frac{V_\phi}{r^2} + \frac{\partial^2 V_\phi}{\partial z^2} \right). \quad (3)$$

Соответственно, уравнение неразрывности для сжимаемой жидкости запишется в виде:

$$\frac{\partial V_r}{\partial r} + \frac{V_r}{r} + \frac{\partial V_z}{\partial z} = 0. \quad (4)$$

Выполнив преобразование (1) и (3) с учетом (4), можно получить зависимость скорости внедрения примеси от гидродинамических параметров расплава в виде следующей системы дифференциальных уравнений:

$$\left\{ \begin{aligned} & V_r \frac{\partial \omega}{\partial r} - \frac{V_r \omega}{r} + V_z \frac{\partial \omega}{\partial z} - \frac{2V_\phi}{r} \frac{\partial V_\phi}{\partial z} = \\ & = v \left(\frac{\partial^2 \omega}{\partial r^2} + \frac{\partial^2 \omega}{\partial z^2} - \frac{\omega}{r^2} + \frac{1}{r^2} \frac{\partial \omega}{\partial r} \right) - g\beta \frac{\partial T}{\partial r}, \\ & V_r \frac{\partial V_\phi}{\partial r} + \frac{V_r V_\phi}{r} + V_z \frac{\partial V_\phi}{\partial z} = \\ & = v \left(\frac{\partial^2 V_\phi}{\partial r^2} + \frac{1}{r} \frac{\partial V_\phi}{\partial r} - \frac{V_\phi}{r^2} + \frac{\partial^2 V_\phi}{\partial z^2} \right), \\ & \omega = \frac{1}{r} \frac{\partial^2 \psi}{\partial r^2} - \frac{1}{r^2} \frac{\partial \psi}{\partial r} + \frac{1}{r} \frac{\partial^2 \psi}{\partial z^2}, \\ & V_r = \frac{1}{r} \frac{\partial \psi}{\partial z}, \\ & V_z = -\frac{1}{r} \frac{\partial \psi}{\partial r}, \end{aligned} \right. \quad (5)$$

где V_r , V_ϕ , V_z – радиальная, тангенциальная и осевая скорости движения расплава, которые определяются частотой вращения тигля; ψ – функция тока расплава; ω – неламинарность расплава; ρ – плотность расплава.

Учитывая, что процесс выращивания проходит при высокой температуре (1420 °C), с учетом уравнения теплопроводности можно записать выражение, определяющее связь между распределением температуры и гидродинамикой расплава в следующем виде:

$$\rho C_p \left(V_r \frac{\partial T}{\partial r} + V_z \frac{\partial T}{\partial z} \right) = \kappa \left\{ \frac{1}{r} \frac{\partial}{\partial r} \left(r \frac{\partial T}{\partial r} \right) + \frac{\partial^2 T}{\partial z^2} \right\}, \quad (6)$$

здесь C_p – удельная изобарная теплоемкость; v – динамическая вязкость; p – давление; β – коэффициент расширения расплава; κ – теплопроводность расплава.

Для решения полученных дифференциальных уравнений введем следующие граничные условия:

1) Для границы расплав–слиток ($0 < r \leq 2.5$ [см], $z = 4.0$ [см])

$$\psi = 0, \quad \omega = \frac{\partial V_r}{\partial z}, \quad V_\phi = r\omega_s, \quad T = T_m.$$

2) Для расплава ($2.5 < r \leq 5.0$ [см], $z = 4.0$ [см])

$$\psi = 0, \quad \omega = \frac{\sigma_T}{\rho v} \frac{\partial T}{\partial r}, \quad \frac{\partial V_\phi}{\partial z} = 0, \quad T = T_{liq}.$$

3) Для стенки тигля ($r = 5.0$ [см], $0 \leq z \leq 4.0$ [см])

$$\psi = 0, \quad \omega = -\frac{\partial V_z}{\partial r}, \quad V_\phi = R_c \omega_c, \quad T = T_{side}.$$

4) Для дна тигля ($0 < r \leq 5.0$ [см], $z = 0$ [см])

$$\psi = 0, \quad \omega = \frac{\partial V_r}{\partial z}, \quad V_\phi = r\omega_c, \quad T = T_{bottom}.$$

5) Для центра слитка ($0 < r \leq 5.0$ [см], $z = 0$ [см])

$$\psi = 0, \quad \frac{\partial \omega}{\partial r} = 0, \quad V_\phi = 0, \quad \frac{\partial T}{\partial r} = 0.$$

Для анализа механизма распределения примеси в расплаве в соответствии с [7] было использовано уравнение конвективной диффузии концентрации примеси, записанное в виде

$$\frac{\partial C_1}{\partial t} + V_r \frac{\partial C_1}{\partial r} + V_z \frac{\partial C_1}{\partial z} = D \left(\frac{\partial^2 C_1}{\partial r^2} + \frac{\partial^2 C_1}{\partial z^2} + \frac{1}{r} \frac{\partial C_1}{\partial r} \right), \quad (7)$$

и уравнение сегрегации в виде

$$C_s = k_0 C_1, \quad (8)$$

где C_s и C_1 – концентрация легирующей примеси в твердой и жидкой фазе; D – коэффициент диффузии; k_0 – коэффициент сегрегации.

Определим граничные условия для решения задачи распределения примеси в кремнии следующим образом:

1) Для границы расплав–слиток ($0 < r \leq 2.5$ [см], $z = 4.0$ [см])

$$\frac{\partial C_1}{\partial z} = \frac{f}{D} (1 - k_0) C_1.$$

2) Для расплава ($2.5 < r \leq 5.0$ [см], $z = 4.0$ [см])

$$\frac{\partial C_1}{\partial z} = 0 \quad \text{или} \quad \frac{\partial C_1}{\partial z} = -\frac{\alpha}{D} C_1$$

3) Для стенки тигля ($r = 5.0$ [см], $0 \leq z \leq 4.0$ [см])

$$\frac{\partial C_1}{\partial z} = 0.$$

4) Для дна тигля ($0 < r \leq 5.0$ [см], $z = 0$ [см])

$$\frac{\partial C_1}{\partial z} = 0.$$

5) Для центра слитка ($r = 0$ [см], $0 \leq z \leq 4.0$ [см])

$$[C_1(0, z) = C_1(1, z)],$$

где α – поток испарения; f – скорость роста слитка.

Результаты моделирования и апробация полученной модели

Для решения поставленной задачи было использовано численное моделирование, которое осуществлялось при помощи двух программных модулей.

Первый модуль «Гидродинамика» решает уравнения Навье–Стокса, уравнения непрерывности и уравнение движения для получения распределения скоростей и температуры расплава кремния (система уравнений (5)). Второй модуль «Примесь» по результатам работы модуля «Гидродинамика» решал уравнения диффузии легирующей примеси и уравнение сегрегации для получения радиального распределения примеси в слитке кремния. Схема взаимодействия программных модулей «Гидродинамика» и «Примесь» показана на рис. 1.

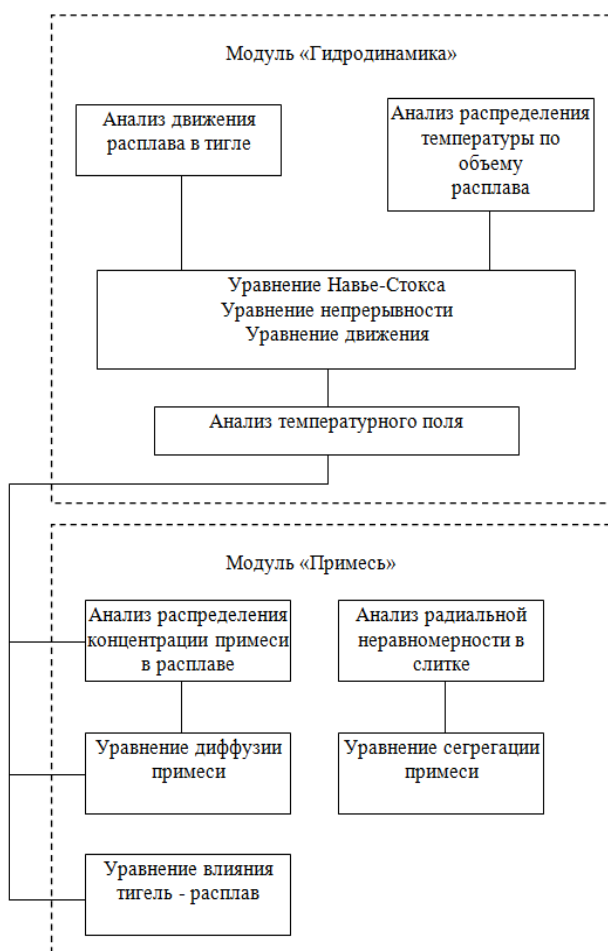


Рис. 1. Схема взаимодействия программных модулей, реализующих вычисления по модели радиальной неравномерности распределения примеси при выращивании слитков кремния методом Чохральского

Для моделирования нами были использованы числовые данные, которые представлены в табл. 1. Для решения дифференциальных уравнений применялся метод конечных разностей. При построении модели была использована сетка, схема которой показана на рис. 2. Распределение концентрации легирующей примеси C_s в слитке определялось как $C_s = k_0 C_1$. В работе учитывалось значение равновесного коэффициента $k_0 = 0,35$ для примеси. В качестве примеси применялась легирующая добавка фосфор (P).

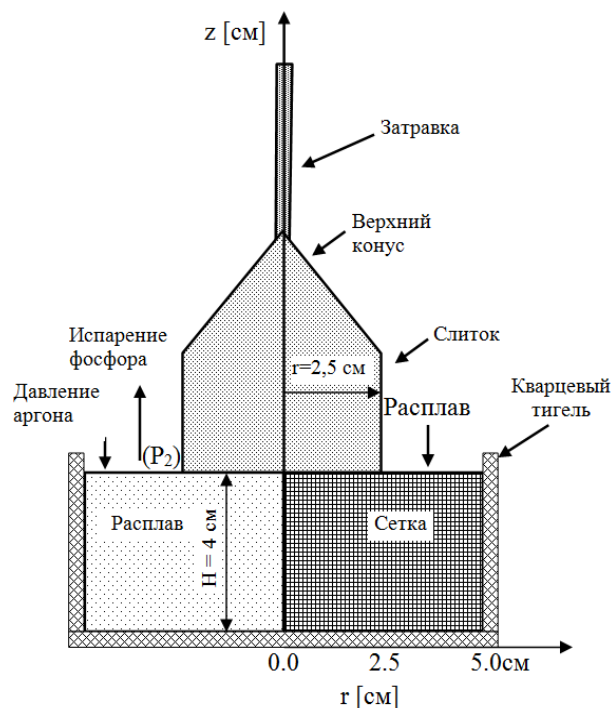


Рис. 2. Блок-схема модели распределения примеси в кремнии в процессе выращивания слитка

Некоторые данные были получены из литературных источников, например коэффициент теплового расширения [8], кинематическая вязкость жидкости [9], плотность жидкости [10] и коэффициент диффузии примеси фосфора в расплаве кремния [11]. Также использовались данные, приведенные в [12].

В процессе расчетов применялась различная дискретизация вычислений и, следовательно, различный шаг расстояния между узлами сетки. По оси z минимальный шаг сетки составлял 0.0025 см для прикристалльной области, в которой происходит внедрение примеси в слиток, и для области вблизи дна тигля. Высоты этих областей выбирались равными 1 см. Шаг сетки и диапазон высоты расплава, для которого производился расчет, приведены в табл. 2. Аналогично было выполнено распределение шага сетки по радиусу слитка (ось r). При этом минимальный шаг использовался для анализа распределения примеси вблизи центра слитка и его внешней границы. Диапазон радиусов и шаг сетки по оси r приведены в табл. 3.

Исходные данные моделирования распределения примеси в кремнии

Символ	Описание	Цифровое значение
C_0	Концентрация жидкой легирующей примеси, перед испарением	5.0×10^{15} атом/см ³
C_s	Концентрация твёрдой легирующей примеси	$C_s(0) = 1$
C_p	Теплоёмкость расплава кремния	0.241 кал/(г×град С)
D	Константа диффузии фосфора в расплаве кремния	5.1×10^{-4} см ² /с
f	Скорость выращивания кристалла	2.33×10^{-3} см/с
H	Высота расплава	4.0 см
k_0	Коэффициент равновесной сегрегации фосфора	0.35
R_c	Радиус тигля	5.0 см
R_s	Радиус выросшего слитка	2.5 см
T_m	Температура плавления кремния	1685 К (1412 °С)
β	Коэффициент расширения расплава	1.43×10^{-4} /К
ν	Кинетическая вязкость расплава	2.789×10^{-3} см ² /с
σ	Поверхностное натяжение	736 дин/см
k	Теплопроводность расплава	0.45 Вт/см К
ρ_l	Плотность расплава	2.533 г/см ³
ρ_s	Плотность закристаллизовавшегося кремния	2.329 г/см ³
ω_c	Частота вращения тигля	7 об/мин
ω_s	Частота вращения слитка	15 об/мин

Таблица 2

Шаг расчетной сетки модели распределения примеси для расплава кремния

Высота расплава [см]	Шаг сетки [см]	Число узлов сетки
0 ÷ 0.10	0.1	1
0.10 ÷ 2.00	0.1000	19
2.00 ÷ 2.50	0.0500	10
2.50 ÷ 3.00	0.1000	5
3.00 ÷ 4.00	0.0250	4
4.00 ÷ 4.50	0.1000	5
4.50 ÷ 4.70	0.0500	4
4.70 ÷ 4.90	0.0250	8
4.90 ÷ 4.95	0.0125	4
4.95 ÷ 5.00	0.0025	20
Всего		80

Таблица 3

Шаг расчетной сетки модели распределения примеси для слитка кремния

Высота расплава [см]	Шаг сетки [см]	Число узлов сетки
4.00 ÷ 3.95	0.0025	20
3.95 ÷ 3.90	0.0125	4
3.90 ÷ 3.70	0.0250	8
3.70 ÷ 3.50	0.0500	4
3.50 ÷ 3.00	0.1000	5
3.00 ÷ 1.00	0.1250	8
1.00 ÷ 0.50	0.1000	5
0.50 ÷ 0.30	0.0500	4
0.30 ÷ 0.10	0.0250	8
0.10 ÷ 0.05	0.0125	4
0.05 ÷ 0.00	0.0025	20
Всего		90

Моделирование движения расплава кремния в тигле показало, что расплав циркулирует от боковой стенки тигля к внутренней части расплава вследствие естественной конвекции, вызванной разницей температур между стенкой тигля и расплавом. В то же время поток расплава ниже области кристаллизации имеет другую форму.

Расплав движется от дна тигля к зоне кристаллизации, и вблизи оси симметрии слитка движение расплава меняет свое направление. Это движение вызвано вращением кристалла и естественной температурной конвекцией в области зоны рециркуляции. На рис. 3 показано распределение температуры в расплаве кремния. Как можно видеть, на распределение температуры влияет естественная конвекция в области зоны рециркуляции. Из-за коэффициента сегрегации, чуть ниже растущего слитка, концентрация легирующей примеси в расплаве для фосфора при скорости потока в 150 см³/мин составляла 5×10^{15} атом/см³. При детальном рассмотрении распределения концентрации фосфора можно увидеть уменьшение концентрации примеси чуть ниже границы расплав-аргон, что вызвано сегрегацией примеси и конвективной диффузией. Расстояние между максимумом и основной массой концентрации расплавленной легирующей примеси определяется как толщина пограничного слоя примеси и приблизительно равна 0,9, 0,6 и 0,4 мм в радиальных положениях 0,0, 1,0 и 2,0 см соответственно. Это значение почти равно толщине гранично-диффузионного слоя примеси, наблюдаемого при передаче рентгеновских изображений [13].

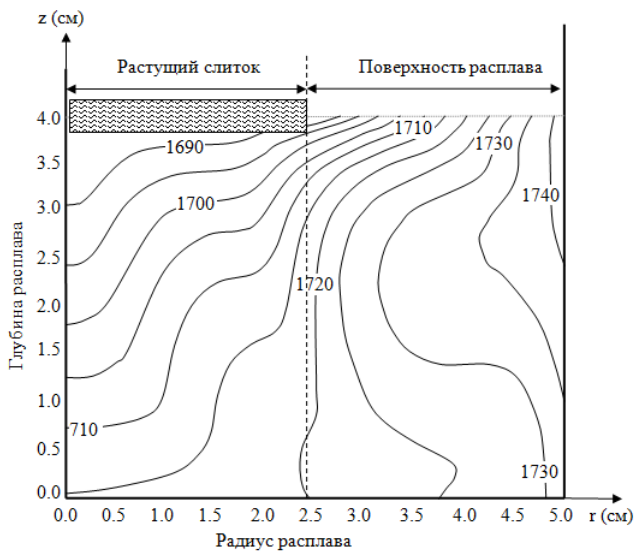


Рис. 3. Симулированное распределение температуры в расплаве Si

Апробация результатов моделирования выполнена путем сравнения экспериментальных измерений и расчетных значений, полученных расчетным путем. Для экспериментальных измерений были выращены слитки монокристаллического кремния в ориентации (100) диаметром 50 мм, из загрузки 1000 г поликристаллического кремния в кварцевом тигле диаметром 100 мм. При этом были использованы следующие параметры технологического процесса выращивания: скорость выращивания – 1,4 мм/мин, частота вращения затравки 15 об/мин, скорость вращения тигля 7 об/мин (в направлении, противоположном направлению вращения затравки). Расплав кремния, был легирован фосфором до уровня $2 \times E15$ атом/см³, что приводит к удельному сопротивлению 2 Ом·м. Слитки, выращенные таким образом, разрезались по оси слитка на шайбы, а удельное сопротивление измерялось с помощью четырехзондового метода [14] с интервалом 1 мм в радиальном направлении. Затем концентрация легирующей примеси была откалибрована с помощью диаграммы Ирвина [15], которая показывает корреляцию между сопротивлением и концентрацией легирующей примеси в кремнии.

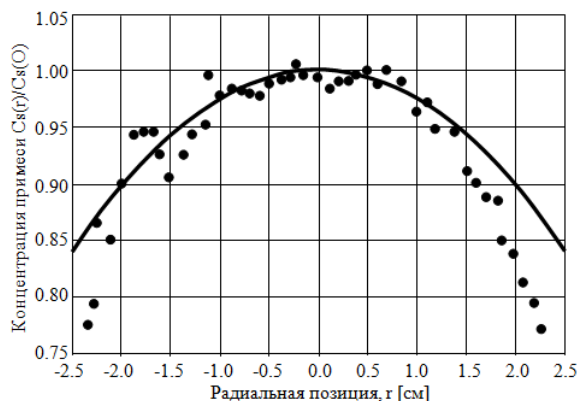


Рис. 4. Измеренное и смоделированное нормированное радиальное распределение концентрации легирующей примеси в кристалле: сплошная линия – рассчитанная кривая, точки – экспериментальные данные

Полученные результаты приведены на рис. 4, где показана концентрация $C_S(r)$, нормированная по концентрации легирующей примеси при $r = 0$, $C_S(0)$.

Можно видеть, что получено хорошее согласование между измеренными и расчетными значениями. Среднеквадратические отклонения между измеренными и расчетными данными составили 2,8 %.

Выводы

1. Усовершенствована математическая модель для виртуального мониторинга радиальной и осевой неравномерности распределения примеси в расплаве и слитке кремния путем учета зависимости скорости внедрения примеси от гидродинамических параметров расплава, отличающаяся тем, что модель использует объединенное решение уравнений Навье-Стокса, уравнения непрерывности, уравнения диффузии примеси и уравнения сегрегации, что позволяет организовать виртуальный мониторинг осевой и радиальной неравномерности распределения примеси и тем самым обеспечить повышение качества кремниевых подложек.

2. В результате апробации модели путем сравнения расчетных и экспериментальных данных получено хорошее согласование между измеренной экспериментальным методом и рассчитанной радиальной концентрацией легирующей примеси в слитках кремния. Отклонение между измеренными и расчетными значениями концентрации не превысило 3%.

3. Усовершенствованную математическую модель можно использовать для построения виртуальных сенсоров, контролирующих распределения примеси в зоне кристаллизации. Решение этой задачи позволит создать информационный модуль, работающий в режиме «Советчик» для уменьшения неравномерности распределения примеси путем корректировки параметров процесса выращивания слитка.

Литература: 1. Pfann W. G. Principles of Zone-Melting / W. G. Pfann. // Transactions of the American Institute of Mining, Metallurgical, and Petroleum Engineers. 1952. №194. P. 747. 2. Haberecht R. R. Semiconductor Silicon / R. R. Haberecht, E. L. Kern. New York: The Electrochemical Society Inc., 1969. 766 p. 3. Sugawara K. A Method to Improve the Radial Impurity Distribution in Czochralski Silicon Crystals / K. Sugawara, H. Tochikubo // Journal of The Electrochemical Society. 1977. №124. P. 951–952. 4. Kim K. M. Computer Simulation of Boron Transport in Magnetic Czochralski Growth of Silicon / K. M. Kim, W. E. Langlois. // Journal of The Electrochemical Society. 1986. №133. P. 2586–2590. 5. Kim K.M. Dopant Segregation in CZ and MCZ Silicon Crystal Growth: A Comparison Between Experiment and Numerical Simulation / K. M. Kim, W. E. Langlois // Journal of The Electrochemical Society. 1989. №136. P. 1137–1142. 6. Kim K.M. Computer Simulation of Oxygen Segregation in CZ/MCZ Silicon Crystals and Comparison with Experimental Results / K. M. Kim, W. E. Langlois // Journal of The Electrochemical Society. 1991. №138. P. 1850–1855. 7. Городцов В. А. Конвективная теплопроводность и диффузия в одномерной гидродинамике / В. А. Городцов // Журнал экспериментальной и теоретической физики. 1999. №116.

C. 1616–1629. **8.** Nakamura S. Thermophysical properties data on molten semiconductors / S. Nakamura, T. Hibiya. // International Journal of Thermophysics. 1992. №13. P. 1061–1084. **9.** Temperature Dependence of the Viscosity of Molten Silicon Measured by the Oscillating Cup Method / [H. Sasaki, E. Tokizaki, X. M. Huang та ін.] // Japanese Journal of Applied Physics. 1995. №34. P. 3432–3436. **10.** Density Variations in Molten Silicon Dependent on Its Thermal History / H. Sasaki, E. Tokizaki, K. Terashima, S. Kimura. // Japanese Journal of Applied Physics. 1994. №33. P. 6078–6081. **11.** Kodaera H. Diffusion Coefficients of Impurities in Silicon Melt / H. Kodaera. // Japanese Journal of Applied Physics. 1963. №2. – P. 212–219. **12.** Physical Properties and critical constants of silicon [Электронный ресурс] // Boston Electronics Corporation – Режим доступа до ресурсу: <http://www.boselec.com/products/matsiphy.html>. **13.** In-situ observation of impurity diffusion boundary layer in silicon Czochralski growth / K. Kakimoto, M. Eguchi, H. Watanabe, T. Hibiya. // Journal of Crystal Growth. 1990. №99. P. 665–669. **14.** Valdes L. B. Resistivity Measurements on Germanium for Transistors / L. B. Valdes // Proceedings of the IRE. 1954. №42. C. 420–427. **15.** Irvin J. C. Resistivity of Bulk Silicon and of Diffused Layers in Silicon / J. C. Irvin. // Bell System Technical Journal. 1962. №14. P. 387–410.

Рецензент: д-р техн. наук, проф. Кривуля Г.Ф.

Поступила в редколлегию 14.01.2016

Оксанич Ирина Григорьевна, канд. техн. наук, доцент кафедры информационно-управляющих систем КрНУ им. М. Остроградского. Научные интересы: информационные технологии управления. Адрес: Украина, 39600, Кременчуг, ул. Першотравнева, 20, тел.: (05366) 30157. Email: alpritchin@ukr.net.

Притчин Алексей Сергеевич, аспирант кафедры компьютерных и информационно-управляющих систем КрНУ им. М. Остроградского. Научные интересы: информационные технологии управления. Адрес: Украина, 39600, Кременчуг, ул. Першотравнева, 20, тел.: (05366) 30157. Email: alpritchin@ukr.net.

Касаткын Артур Владимирович, аспирант кафедры компьютерных и информационно-управляющих систем КрНУ им. М. Остроградского. Научные интересы: информационные технологии управления. Адрес: Украина, 39600, Кременчуг, ул. Першотравнева, 20, тел.: (05366) 30157. Email: alpritchin@ukr.net.



ТЕХНОЛОГИЯ ПОВЫШЕНИЯ БЕЗОПАСНОСТИ ИНФОРМАЦИОННЫХ РЕСУРСОВ НА ОСНОВЕ ИСПОЛЬЗОВАНИЯ ФУНКЦИОНАЛЬНОГО ПРЕОБРАЗОВАНИЯ ПРИ КОСВЕННОМ СТЕГАНОГРАФИЧЕСКОМ ВСТРАИВАНИИ

ЮДИН А.К., БАРАННИК В.В., ФРОЛОВ О.В.

Рассматривается возможность повышения безопасности информационных ресурсов на основе использования методов компьютерной стеганографии. Проводится сравнительный анализ методов непосредственного и косвенного стеганографического встраивания информации в изображение контейнер. Определяются основные показатели качества функционирования систем косвенного стеганографического встраивания. Проводится сравнительный анализ наиболее распространенных существующих стеганографических методов. Для устранения выявленных недостатков существующих стеганографических систем сформулирован подход, основанный на синтезировании функционала для использования при встраивании структурной избыточности изображения.

Введение

Одним из возможных способов повышения безопасности информационных ресурсов в инфокоммуникационных каналах является использование методов компьютерной стеганографии. На сегодняшний день наибольшей популярностью пользуются методы, позволяющие встраивать скрываемую информацию в изображение-контейнер. Данный факт обусловлен широким распространением цифровых изображений различных форматов.

Существующие стеганографические методы встраивания информации в изображение-контейнер включают алгоритмы непосредственного стеганографического встраивания и алгоритмы косвенного стеганографического встраивания. Непосредственное встраивание информации в изображение реализуется путем замены бита контейнера на бит скрываемого сообщения. Алгоритмы, основанные на данном подходе, не обеспечивают в полной мере системных требований относительно стойкости встроенных данных к активным атакам злоумышленника. Это обосновано в первую очередь тем, что методы непосредственного стеганографического встраивания используют визуаль-

ную избыточность изображения для скрытия данных. При этом встроенные данные неустойчивы к атакам с пережатием, которые направлены на устранение потенциально опасной психовизуальной избыточности.

В отличие от методов непосредственного встраивания, стеганографическое встраивание бита скрываемого сообщения в косвенных методах осуществляется путем создания зависимости между некоторыми параметрами изображения-контейнера согласно определенному алгоритму. При этом обратное стеганографическое преобразование осуществляется путем изъятия некоторой оценки встроенных данных.

Отсюда предлагается направление для повышения безопасности информационных ресурсов на основе использования методов косвенного стеганографического встраивания.

Анализ существующих методов косвенного стеганографического встраивания

Для успешного сравнительного анализа существующих методов косвенного встраивания рассмотрим стеганографические показатели качества:

1. Относительная стеганографическая емкость $w_{отн}$. Данный показатель используется для оценки эффективности стеганографической системы по удельному объему $w_{встр}$ встраиваемой информации относительно объема $W_{исх}$ изображения-контейнера. Величина $w_{отн}$ относительной стеганографической емкости системы определяется на основе следующей

$$\text{формулы: } w_{отн} = \frac{w_{встр}}{W_{исх}}.$$

2. Вероятность $P_{из}$ безошибочного изъятия встроенных данных для авторизированного доступа:

$$P_{из} = \frac{w_{из}}{w_{встр}},$$

где $w_{встр}$ – объем встраиваемой информации, бит; $w_{из}$ – объем безошибочно изъятых информации, бит.

3. Пиковое отношение сигнал-шум h изображения со встроенными данными. Этот показатель характеризует стеганографическую систему с позиции устойчивости к визуальным атакам злоумышленника:

$$h = 20 \log_{10}(255/СКО), \text{ дБ},$$

где СКО – среднеквадратическое отклонение изображения со встроенными данными относительно изображения-контейнера.

4. Вероятность $P_{уст}$ безошибочного изъятия встроенных данных в условиях применения злоумышленником активных атак определяется как отношение объема $w_{встр}$ встроенных данных к объему $w'_{из}$

безошибочно изъятой информации в условиях активных атак:

$$P_{уст} = \frac{w'_{из}}{w_{встр}}$$

и составляет 100%.

5. Пиковое отношение сигнал-шум $h_{авт}$ изображения, полученного в процессе обратного стеганографического преобразования при авторизированном доступе:

$$h_{авт} = 20 \log_{10} \left(255 / \sqrt{\sum_{i=1}^{z_{строк}} \sum_{j=1}^{z_{столб}} (a_{i,j} - a'_{i,j})^2} / z_{строк} z_{столб} \right)$$

где $a_{i,j}$, $a'_{i,j}$ – элементы соответственно исходного и реконструированного при авторизированном доступе изображений; $z_{строк}$, $z_{столб}$ – размер изображения-контейнера.

Методы косвенного стеганографического встраивания условно разделены на две базовые группы, а именно:

1. Методы, для которых требуется прототип изображения-контейнера при стеганографическом изъятии.
2. Методы, для которых не требуется прототипа изображения-контейнера.

Рассмотрим принципы функционирования наиболее распространенных методов, которые используют для косвенного стеганографического изъятия прототип изображения-контейнера.

1. Метод Подильчука. Данный метод предусматривает вычисление порога коэффициента ДКП изображения-контейнера на основе его позиции в матрице. При встраивании анализируются вычисленные пороги. Если значение коэффициента меньше порога, тогда он не изменяется. В противном случае к коэффициенту прибавляется произведение порога и значение элемента скрываемого сообщения. Изъятие встроенной информации осуществляется путем сравнения коэффициентов ДКП стеганограммы и коэффициентов ДКП прототипа исходного изображения-контейнера.

2. Метод Гао. При стеганографическом преобразовании на первом этапе выполняется классификация блоков по шести категориям в зависимости от степени гладкости и наличия контуров. Для каждого блока на основе ключевого правила вычисляются коэффициенты чувствительности к шуму. На следующем этапе блоки упорядочиваются в соответствии с полученными коэффициентами. Энергия встраиваемого элемента определяется этими коэффициентами. Для изъятия встроенных данных выполняется вычитание прототипа изображения-контейнера из принятой стеганограммы и применяются статистические методы проверки гипотез.

3. Алгоритм Кокса. В качестве встраиваемой информации для данного метода используются последовательности вещественных чисел с нулевым средним и единичной дисперсией. Для встраивания применяются несколько АС-коэффициентов ДКП изображения-контейнера с наибольшей энергией. При встраивании осуществляется модификация АС-коэффициентов в соответствии с ключевым правилом встраивания. При изъятии осуществляются обратные операции: определяются коэффициенты ДКП стеганограммы и прототипа изображения-контейнера, находится разность между коэффициентами наибольшей величины.

В отличие от методов, использующих прототип изображения-контейнера при изъятии, методы без учета прототипа обеспечивают изъятие встроенных данных «вслепую». Другими словами, для реализации обратного стеганографического встраивания на основе таких методов не требуется наличие исходного изображения-контейнера.

Среди указанных методов можно выделить следующие:

1. Метод относительной замены величин ДКП (метод Коха и Жао). Одним из наиболее распространенных на сегодня методов. В его алгоритме реализовано разбиение изображения на блоки 8*8 пикселей для применения к каждому из них ДКП. В результате данного преобразования получается матрица 8*8 коэффициентов ДКП. Каждый блок используется для скрытия одного бита данных. Для обеих сторон при организации секретного канала выбираются два конкретных коэффициента ДКП с определенными координатами в массиве коэффициентов. Непосредственно скрытие начинается со случайного выбора блока изображения, предназначенного для кодирования бита данных. Встраивание происходит такой модификацией коэффициентов, чтобы при передаче «0» их разница превышала некоторую положительную величину, а для «1» эта разница делается меньшей по сравнению с некоторой отрицательной величиной. Таким образом, первичное изображение модифицируется за счет внесения изменения в коэффициенты ДКП. После соответствующей коррекции коэффициентов проводится обратное дискретное косинусное преобразование.

2. Метод модификации яркости (метод Куттера-Джордана-Боссена). Встраивание реализуется в канал синего цвета RGB изображения. Цвет был выбран из-за низкой чувствительности человека к его изменению. Секретный бит M_i встраивается в канал синего цвета путем модификации яркости

$$\lambda_{x,y} = 0.29890 \cdot R_{x,y} + 0.58662 \cdot G_{x,y} + 0.11448 \cdot B_{x,y} :$$

$$B'_{x,y} = B_{x,y} - v \cdot \lambda_{x,y} \quad \text{при} \quad m_i = 0$$

$$\text{и} \quad B'_{x,y} = B_{x,y} - v \cdot \lambda_{x,y} \quad \text{при} \quad m_i = 1,$$

где v - величина, которая определяет энергию встраиваемого сигнала, прямо пропорциональна устойчивости встроеной информации к искажениям.

Для извлечения секретного бита получателю необходимо выполнить предсказание значения первичного не модифицированного пикселя, используя значения соседних пикселей. Авторы метода использовали «крест» пикселей размером 7×7 .

3. Метода Бенгама-Мемона-Эо-Юнг. Встраивание осуществляется в спектральные коэффициенты изображения-контейнера путем их модификации. Для этого в спектральной области выбираются три коэффициента ДКП, что позволяет уменьшить визуальные искажения. Для встраивания «0» эти коэффициенты изменятся таким образом, чтобы третий коэффициент стал меньше любого из двух первых. Если необходимо скрыть «1», он делается большим, чем первый и второй коэффициенты. Использование трех коэффициентов ДКП вместо двух уменьшает искажения, которые вносятся в результате встраивания, скрываемого сообщением.

Существующие методы косвенного стеганографического встраивания не обеспечивают в полной мере требований относительно безопасного скрытия данных в изображении. Это обусловлено следующими недостатками:

- низкое значение устойчивости стеганограммы к визуальным атакам злоумышленника. Данный недостаток обусловлен тем, что встраивание скрываемой информации достигается путем модификации элементов представления стеганограммы. Это сопровождается внесением визуальных искажений в изображение и ухудшением его качества. В случае наличия у злоумышленника исходного изображения-контейнера может быть выявлен факт наличия скрытого встраивания в стеганограмме;

- низкая устойчивость встроенных данных к активным атакам злоумышленника. Среди таких атак наиболее распространенными являются компрессионные атаки. Они направлены на устранение психовизуальной избыточности, которая также используется для косвенного стеганографического встраивания информации. Применяя данные атаки, противник способен безвозвратно разрушить встроеное сообщение;

- неудовлетворительное значение стеганографической емкости. Существующие методы встраивания не обеспечивают требуемого объема встраиваемой информации. Данный недостаток обусловлен тем, что увеличение объема встраивания сопровождается увеличением числа модифицированных элементов и как следствие увеличением вносимых искажений в изображение;

- необходимость наличия на приемной стороне прототипа исходного изображения-контейнера для однозначного изъятия встроеной информации.

Существующие недостатки обусловлены тем, что методы используют для косвенного встраивания психовизуальную избыточность изображения.

Основная часть

Для устранения выявленных недостатков косвенного стеганографического встраивания предлагается разработать подход, который позволит использовать для скрытого встраивания структурную избыточность изображения-контейнера. В качестве такого подхода предлагается синтезировать функциональное преобразование $f(\bullet)$ для элементов изображения-контейнера, которое должно обеспечить следующие требования:

1. Функциональное преобразование должно обеспечить взаимоднозначное кодирование $f(\bullet)$ и декодирование $f^{-1}(\bullet)$ массива A изображения-контейнера при наличии служебной информации Ψ , т.е.

$$C = f(A, \Psi), A' = f^{-1}(C, \Psi) \text{ и } A' = A.$$

Здесь A' – массив, восстановленный в результате обратного функционального преобразования $f^{-1}(C)$; Ψ – служебная информация; C – массив, полученный в результате выполнения прямого функционального преобразования.

2. В результате функционального преобразования массива A должна формироваться кодограмма C , которая состоит из двух частей:

- служебной составляющей, содержащей служебные данные Ψ ;

- информационной составляющей, содержащей кодовое представление массива A .

3. Значения реконструированных массивов A' и A'' не должны меняться в случае формирования кода при различных значениях служебной информации (Ψ и Ψ'), т.е. $A' = f^{-1}(C, \Psi) = f^{-1}(C', \Psi') = A''$, где A' – массив, реконструированный на основе кода, сформированного с учетом служебных данных Ψ ;

A'' – массив, реконструированный на основе кода, сформированного с учетом модифицированных служебных данных Ψ' ;

C – кодограмма, полученная с учетом служебных данных Ψ ;

C' – кодограмма, полученная с учетом служебных данных Ψ' .

Предлагается использовать данное свойство для косвенного стеганографического встраивания.

Тогда процесс встраивания будет включать намеренное изменение служебной информации Ψ на основе ключевого условия. Сформированная кодограмма C' , содержащая модифицированные служебные данные Ψ' , передается по каналу данных.

При этом на приемной стороне авторизованному пользователю известно условие косвенного встраивания, т.е. механизма модификации исходной служебной информации Ψ . В этом случае процесс стеганографического изъятия будет осуществляться путем

анализа значений исходной Ψ и измененной Ψ' служебной информации.

Прямое косвенное стеганографическое преобразование включает следующие этапы:

1. Формирование вектора служебных данных Ψ для массива $A''(2)$ изображения-контейнера.

2. Второй этап предусматривает модификацию вектора служебных данных Ψ с учетом встраиваемого элемента b_ξ на основе ключевого условия: $\Psi' = \Psi + b_\xi$. Здесь b_ξ – элемент скрываемого сообщения $B = \{b_1; \dots; b_\xi; \dots; b_v\}$, $\xi = \overline{1, v}$.

3. Функциональное преобразование массива A с учетом модифицированного вектора служебных данных Ψ' по правилу $f(A)$, т.е. $C = f(A, \Psi')$, где C – сформированное значение кодограммы.

Полученная кодограмма, содержащая в себе информационную составляющую C и служебную составляющую Ψ' , передается в канал передачи данных, где может подвергаться атакующим воздействиям.

Обратное косвенное стеганографическое преобразование осуществляется по биполярному принципу для авторизованного и неавторизованного пользователя.

При неавторизованном доступе по правилу $f^{(-1)}(\bullet)$ осуществляется реконструкция исходного массива изображения-контейнера: $A'' = f^{(-1)}(C; \Psi')$.

Здесь A'' – массив исходного изображения, полученный в результате неавторизованного доступа.

Наоборот, обратное косвенное стеганографическое преобразование для авторизованного пользователя осуществляется с учетом ключевого условия изъятия и содержит следующие этапы:

1. На первом этапе по правилу $f^{(-1)}(\bullet)$ реконструируется массив A' исходного изображения контейнера:

$$A' = f^{(-1)}(C; \Psi')$$

Здесь C – принятая кодограмма, сформированная на передающей стороне с учетом модифицированных служебных данных Ψ' .

2. На втором этапе для реконструированного массива A' по ключевому правилу осуществляется формирование исходного вектора служебных данных Ψ .

3. Третий этап включает косвенное изъятие встроенного элемента b'_ξ скрываемого сообщения $B' = \{b'_1; \dots; b'_\xi; \dots; b'_v\}$ на основе ключевого условия изъятия при анализе восстановленного Ψ и полученного Ψ' векторов служебных данных: $b'_\xi = \Psi' - \Psi$.

Выводы

Рассмотрена возможность повышения безопасности информационных ресурсов в инфокоммуникационных системах на основе использования методов косвенного стеганографического встраивания.

Приведены основные показатели качества функционирования систем косвенного стеганографического встраивания. Проведен сравнительный анализ наиболее распространенных косвенных стеганографических методов. Определены основные недостатки функционирования таких систем.

Для устранения выявленных недостатков предложен подход, основанный на использовании при косвенном стеганографическом встраивании структурной избыточности представления изображения-контейнера.

Сформулированы требования к синтезированному функционалу. Представлена схема косвенного стеганографического преобразования на основе использования синтезированного функционального преобразования.

- Литература:** 1. Грибунин В.Г., Оков И.Н., Туринцев И.В. Цифровая стеганография М.: Солон-Пресс, 2002. 272 с. 2. Конахович Г.Ф., Пузыренко А.Ю. Компьютерная стеганография. Теория и практика. К.: МК-Пресс, 2006. 288с. 3. Тарасов Д.О., Мельник А.С., Голобородько М.М. Класифікація та аналіз безкоштовних програмних засобів стеганографії // Інформаційні системи та мережі. Вісник НУ “Львівська політехніка”. 2010. № 673. С. 365-374. 4. Защита от компьютерного терроризма / А. В. Соколов, О. М. Степанюк // Справочное пособие. БВХ-Петербург: Арлит. 2002. 496 с. 5. Жилкин М.Ю. Стегоанализ графических данных в различных форматах / М.Ю. Жилкин // Доклады ТУСУРа, 2008. №2 (18). Ч. 1. С.63-64. 6. Задирака В.К. Новые подходы к разработке алгоритмов скрытия информации / В.К. Задирака, Л.Л. Никитенко // Штучний інтелект. 2008. №4. С.353-357. 7. Михайличенко О.В. Применени стеганографических методов сокрытия информации в неподвижных изображениях / О.В. Михайличенко, А.Г. Коробейникова, С.Ю. Каменева // Труды международных научно-технических конференций «Интеллектуальные системы (IEEE AIS'06) и «Интеллектуальные САПР (CAD-2006)». М.: Физмалит, 2006. Т.2. С.511-515.

Поступила в редколлегию 17.01.2016

Рецензент: д-р техн. наук, проф. Безрук В.М.

Баранник Владимир Викторович, д-р техн. наук, профессор, начальник кафедры автоматизированных систем управления, Харьковский университет Воздушных Сил им. И. Кожедуба. Научные интересы: кодирование и защита информации для передачи в телекоммуникационных системах. Адрес: Украина, 61000, Харьков, ул. Сумская, 77/79. E-mail: barannik_v_v@mail.ru.

Юдин Александр Константинович, д-р техн. наук, профессор, директор института компьютерных информационных технологий Национального авиационного университета. Адрес: Украина, 01000, Киев, пр.Космонавта Комарова, 1.

Фролов Олег Владимирович, соискатель Национального авиационного университета. Адрес: Украина, 01000, Киев, пр.Космонавта Комарова, 1.



МЕТОД ШТРАФНЫХ ФУНКЦИЙ ДЛЯ РЕШЕНИЯ ЗАДАЧ ОПТИМИЗАЦИИ НА ПОЛИЭДРАЛЬНО-СФЕРИЧЕСКИХ КОМБИНАТОРНЫХ МНОЖЕСТВАХ

ПИЧУГИНА О.С., ЯКОВЛЕВ В.С.

Предлагается новый подход к решению задач оптимизации на вписанных в сферу евклидовых комбинаторных множествах, основанный на применении функциональных представлений дискретных множеств и выпуклых продолжений с них в \square^n в методе штрафных функций.

Ключевые слова: комбинаторная оптимизация, евклидовое комбинаторное множество, выпуклое продолжение функций, непрерывное представление множеств, метод штрафных функций.

Key words: combinatorial optimization, the euclidean combinatorial set, a convex extension of functions, a continuous representation of sets, the penalty method.

Введение

Задачи дискретной оптимизации традиционно считаются сложными [1-4], поэтому точные методы имеют, как правило, теоретический интерес, в то время как многочисленные практические задачи, сводящиеся к дискретным, зачастую разрешимы лишь эвристически [5]. Задачи комбинаторной оптимизации как подкласс дискретных являются более перспективными, поскольку позволяют разрабатывать методы, учитывающие алгебро-топологическую структуру этих множеств и функций на них [6-8]. Особый интерес представляют подходы, основанные на применении к комбинаторным задачам методов непрерывной, в том числе выпуклой, оптимизации [2-3, 9-16]. В них, в свою очередь, выделяются релаксационные методы, основанные на непрерывных переформулировках комбинаторных задач. Последние, как правило, предусматривают существенное увеличение размерности задачи. Известны также случаи, когда исходные комбинаторные задачи сводятся к оптимизации выпуклых функций, что дает возможность создания как эффективных приближенных алгоритмов, так и точных методов типа ветвей и границ, использующих для оценок выпуклые релаксации.

В данной статье рассматривается именно такой класс задач оптимизации на вписанных в сферу евклидовых комбинаторных множествах (далее полиэдрально-сферических множествах), позволяющих непрерыв-

ные постановки с выпуклыми целевыми функциями в пространстве исходной размерности.

Цель исследования – разработка непрерывного подхода к решению задач оптимизации на полиэдрально-сферических множествах, основанного на использовании их аналитических непрерывных представлений и продолжений целевых функций с них в надмножества \square^n , включая выпуклые продолжения [17-19].

Основные задания:

1. Ввести понятия и дать классификацию функциональных представлений евклидовых комбинаторных множеств и продолжений функций с них.
2. Для безусловных задач комбинаторной оптимизации на полиэдрально-сферических ЕКМ представить эквивалентные непрерывные формулировки в терминах строгих функциональных представлений допустимой области.
3. Обосновать возможность построения выпуклых, дифференцируемых и строгих продолжений целевой функции с полиэдрально-сферических множеств в их выпуклые надмножества и целесообразности их использования в оптимизационном методе.
4. Провести сравнительный анализ различных способов выпуклых продолжений штрафных функций.
5. Описать и обосновать метод решения задач оптимизации на полиэдрально-сферических множествах с использованием метода штрафных функций.

Оптимизационные задачи на полиэдрально-сферических комбинаторных множествах охватывает широкий класс задач евклидовой комбинаторной оптимизации [6,7], как безусловных и условных, к ним сводящихся. Это задачи оптимизации на общем множестве перестановок [4,6,7,20], множестве B_n n -мерных булевых векторов [10,13-16], отдельных классах размещений и сочетаний [7], а также на их подмножествах, декартовых произведениях и некоторых других композиционных образах [21]. Примером являются множество полиперестановок [7,21], такие подмножества B_n как [22]:

$$B_n(k) = \left\{ x : \sum_{i=1}^n x_i = k \right\} \text{ и } B_n(0, k) = \left\{ x : \sum_{i=1}^n x_i \leq k \right\}.$$

Итак, рассматриваемый класс задач охватывает безусловные булевы задачи, известные разнообразием применений и множеством методов решения [3,14-16,22-25,27]. Перечислим некоторые задачи на подмножествах B_n . На $B_n(k)$ формулируются: задача бисекции графа [0] (the graph bisection problem); задача о наиболее плотном k -подграфе (the densest k -subgraph problem) [23], обобщающая задачу о k -кластерах (k -cluster problem) [220] и целый ряд классических задач теории графов, таких как задача о максимальной клике [1]. Еще одним широко исполь-

зующимся подмножеством B_n является множество перестановочных матриц [24], оптимизационные задачи на котором включают линейную и квадратичную задачи о назначениях [25] и множество других задач. Задачи на перестановочных матрицах зачастую позволяют упрощенную формулировку на перестановочном множестве E_n [4,6], которое также относится к классу полиэдрально-сферических.

1. Постановка задачи и необходимые определения

Рассмотрим безусловную задачу на полиэдрально-сферическом множестве:

$$z^* = \min_E f(x), x^* = \arg \min_E f(x), \quad (1)$$

$$E = P \cap S_r(a), \quad (2)$$

где

$$P = \text{conv } E = \{x \in \square^n : Ax \leq b\} - \text{многогранник}, \quad (3)$$

$$S_r(a) = \{x \in \square^n : (x-a)^2 = r^2\} - \text{сфера}. \quad (4)$$

Условия (2), (3) означают, что E – вершинно расположено:

$$E = \text{vert } P \quad (5)$$

и конечно. Будем также считать, что E – не вырожденное в точку множество, т.е.

$$1 < N = |E| < \infty. \quad (6)$$

Напомним, что евклидовым комбинаторным множеством (ЕКМ) [8] называется произвольный числовой набор объектов комбинаторной природы, элементы которого отличаются составом либо порядком следования элементов. Иначе говоря, ЕКМ – это комбинаторное множество, позволяющее погружение в \square^n , при этом полученный образ также является ЕКМ [7].

Итак, поставленная задача (1)-(2) – безусловная задача на конечном полиэдрально-сферическом евклидовом комбинаторном множестве.

Функциональным назовем представление множества E при помощи функциональных зависимостей вида:

$$f_j(x) = 0, j \in J_{m'}, \quad (7)$$

$$f_j(x) \leq 0, j \in J_m \setminus J_{m'}, \quad (8)$$

где $J_m = \{1, \dots, m\}$.

В обозначениях $m = m' + m''$ представление (7)-(8) назовем:

– строгим представлением E в случае $m = m'$,

иначе – нестрогим;

– непрерывным представлением E , если

$$f_j(x), j \in J_m - \text{непрерывные}; \quad (9)$$

– выпуклым представлением E , когда

$$f_j(x), j \in J_m - \text{выпуклые}$$

на некотором выпуклом множестве, содержащем E .

Так, система (3), (4) задает непрерывное, нестрогое, выпуклое функциональное представление E , называемое, в силу (2), полиэдрально-сферическим.

Предположим, что для E известно также строгое представление, т.е. система (7), (8) приобретает вид:

$$f_j(x) = 0, j \in J_m. \quad (10)$$

Продолжением функции $f(x)$ с E в $E' \supseteq E$ называется функция $F(x)$, определенная на E' и совпадающая с $f(x)$ на E :

$$F(x) = f(x). \quad (11)$$

В зависимости от типа функции $F(x)$ такие продолжения могут быть непрерывными, выпуклыми, дифференцируемыми и т.п.

Продолжение функции $f(x)$ с E в $E' \supseteq E$ называется строгим, если

$$F(x) = f(x) \Leftrightarrow x \in E. \quad (12)$$

Продолжение функции $f(x)$ с E в \square^n называется ее продолжением с E или просто продолжением $f(x)$.

2. Анализ литературных данных

Основные направления современной комбинаторной оптимизации условно делятся на:

– комбинаторные методы, такие как методы ветвей и границ, отсечений, ветвей и отсечений, эвристики [1-3,5,16];

– непрерывные методы, среди которых выделяются [2,3,26]:

релаксационные подходы, в том числе основанные на полу-определенных релаксациях [22];

подходы, использующие эквивалентные непрерывные переформулировки исходной дискретной задачи [26].

Все эти методы используют в большей или меньшей степени свойства конкретного комбинаторного множества и функций, например, вогнутых и выпуклых, на нем.

Остановимся далее на непрерывных подходах, для применения которых комбинаторное множество должно быть представимо в непрерывных переменных. Как правило, при этой трансформации размерность

пространства увеличивается [3,16,22,26]. Наличие же функционального представления (7)-(8) допустимого множества означает возможность непрерывной постановки комбинаторной задачи в исходном пространстве и открывает перспективы создания новых методов ее решения.

Так, в булевой оптимизации широко применяется следующий вид булевых множеств $B_n = \{0,1\}^n$ [10, 14-15,26], $B'_n = \{-1,1\}^n$ [10,22,27]:

$$B_n = \left\{ x \in \square^n : f_i(x) = x_i^2 - x_i = 0, i \in J_n \right\}, \quad (13)$$

$$B'_n = \left\{ x \in \square^n : f_j(x) = x_j^2 - 1 = 0, j \in J_n \right\}. \quad (14)$$

В нашей терминологии (13),(14) – строгие функциональные представления (10) булевых множеств, где $m = n$. Поэтому множества B_n, B'_n в виде представлений (13),(14) выбраны для иллюстрации излагаемого материала и примеров.

Отличительной особенностью вершинно расположенных множеств вообще и полиэдрально-сферических в частности является возможность формирования выпуклых, вогнутых и дифференцируемых продолжений с них [17], что находит применение как в непрерывных, так и в комбинаторных методах.

Например, для безусловной булевой задачи

$$f(x) \rightarrow \min_{B_n}, f(x) \in C^2(\square^n), \quad (15)$$

применимы следующие непрерывные методы, использующие непрерывные функциональные представления множества B_n и продолжений с него:

– вогнутое программирование, при котором функциональное представление (13) аккумулируется в одно ограничение, в результате чего (15) сводится к минимизации вогнутого продолжения $f(x)$ на единичном гиперкубе [26];

– методы штрафных функций, в которых используется представление (13) и, с его помощью, задача (15) сводится к серии выпуклых задач оптимизации квадратичной штрафной функции, являющейся выпуклым продолжением сглаживания $f(x)$ [14-15].

– полиэдрально-сферический метод [11], основанный на использовании представлений (2),(13) и сводящий (15) к серии задач минимизации выпуклого продолжения $f(x)$ на единичных гиперкубах и сферах различной размерности.

3. Мотивация к методу. Примеры

Рассматриваемые полиэдрально-сферические ЕКМ обладают важными особенностями, положенными в основу предлагаемого метода решения (1), (2) (см. п.3.4). А именно, это существование строгих функциональных представлений, а также возможность построения как выпуклых, так и строгих их продолжений. Кратко обоснуем это.

Существование строгих представлений множеств типа (2) показано на примере B_n, B'_n (см. (13)-(14)). Для произвольного такого множества также может быть найдено представление $E = S \cap S_r(a)$, где S – сглаживающая поверхность многогранника (3) выпуклая поверхность, вписанная в $S_r(a)$: $E \subset S$. В рамках данной статьи задача построения строгих представлений не ставится. Предполагается, что она уже решена и (10) известно.

Существование выпуклых продолжений $f(x)$ с E в произвольное выпуклое его надмножество основано на следующих фактах: вписанность конечного множества E в сферу означает выполнение условия (5), а, как известно [17], для произвольной функции, определенной на таком множестве, существует выпуклое дифференцируемое продолжение с E .

Поэтому далее полагаем, что $f(x)$ – выпукла:

$$f(x) \in C^1(\square^n), \quad (16)$$

иначе переходим к рассмотрению ее выпуклого дифференцируемого продолжения [17].

Замечание 1. Если целевая функция (1):

$$f(x) \in C^2(\square^n), \quad (17)$$

то, согласно [19], существует ее продолжение вида:

$$F(x) = f(x) + M((x-a)^2 - r^2), \quad (18)$$

$$F(x) \in C^2(\square^n) \quad (19)$$

с E в произвольный компакт $E' \supset E$.

Будем полагать далее, что (17), выполнено.

Если к тому же выполнено

$$\frac{\partial^2 f(x)}{\partial x_i \cdot \partial x_j} < \infty, i, j \in J_n, \quad (20)$$

то существует продолжение вида (18)-(19) функции $f(x)$ с E в \square^n .

Покажем, что существуют также строгие продолжения $f(x)$ с множества E . Построим их при помощи строгого представления (10) этого множества, которое, по предположению, известно.

Выберем произвольный вектор $\bar{\lambda} = (\lambda_i)_{i \in J_m} > 0$ и построим функцию вида:

$$F'(x, \bar{\lambda}) = f(x) + \sum_{i=1}^m \lambda_i f_i^2(x). \quad (21)$$

Нетрудно видеть, что (21) – продолжение $f(x)$. К тому же, в соответствии с определением строгого представления: $x \in E \Leftrightarrow \left\{ f_i(x) = 0 \right\}_{i \in J_m}$. А это означает:

$$x \in E \Leftrightarrow \sum_{i=1}^m \lambda_i f_i^2(x) = 0 \Leftrightarrow F'(x, \bar{\lambda}) = f(x) + \sum_{i=1}^m \lambda_i f_i^2(x).$$

Следовательно, для $F'(x, \bar{\lambda})$ выполнено (12), таким образом, (21) – строгое продолжение $f(x)$ с E .

Не ограничивая общности, можно считать, что составляющие строгого представления (10) – дифференцируемые:

$$f_j(x) \in C(\square^n), j \in J_m, \quad (22)$$

иначе переходим к рассмотрению их дифференцируемых продолжений с E .

В совокупности допущения (16),(22) обосновывают существование дифференцируемых (а в случае (17) и дважды непрерывно-дифференцируемых) строгих продолжений $f(x)$.

Итак, показано, что для произвольного полиэдрально-сферического конечного множества существуют строгие функциональные представления. К тому же для любой целевой функции (1), заданной на E , существуют строгие, выпуклые, дифференцируемые (а при выполнении (17) – дважды непрерывно-дифференцируемые) продолжения с E . А это, в свою очередь, обосновывает возможность применения к ним непрерывной, а в отдельных случаях и выпуклой, оптимизации в исходном пространстве.

3.1. Непрерывные формулировки исходной задачи

Приведем некоторые непрерывные постановки задачи (1),(2),(10):

1) Первая непрерывная формулировка этой задачи состоит в использовании полиэдрально-сферического представления (2)-(4) и замены условия дискретности $x \in E$ условиями принадлежности континуальным множествам (3) ($x \in P$) и (4) ($x \in S_T(a)$).

2) Еще одна непрерывная постановка в виде условной нелинейной задачи состоит в применении строгого представления E и замене в (1) условия $x \in E$ на (10) (нелинейность системы объясняется тем, что линейная система задавала бы выпуклое множество, что противоречит дискретности и невырожденности (6)).

3) Наконец, условия (10) могут быть интегрированы в функцию Лагранжа, в результате чего исходная задача эквивалентно формулируется как безусловная задача ее минимизации:

$$\phi(x, \bar{\lambda}) = f(x) + \sum_{i=1}^m \lambda_i f_i(x) \rightarrow \min. \quad (23)$$

При этом новая целевая функция по-прежнему непрерывно-дифференцируема в силу (16),(22). Задача (23) решается в пространстве \square^{n+m} .

4) Следующая эквивалентная непрерывная формулировка задачи (1),(2) основана на использовании строгого функционального представления в штрафных функциях (10). Например, для квадратичной штрафной функции она выглядит следующим образом:

$$\varphi(x, \lambda) = f(x) + \lambda \sum_{i=1}^m f_i^2(x) \xrightarrow{\lambda \rightarrow \infty} \min, \quad (24)$$

где $\lambda > 0$ – штрафной множитель.

Эквивалентность задач (1),(2) и (24) основана на следующих свойствах:

1. $\forall x \in R^n, \lambda > 0 \varphi(x, \lambda) \geq f(x)$;
2. $\forall \lambda > 0 \varphi(x, \lambda) = f(x) \Leftrightarrow x \in E$, т.е. (24) – строгое продолжение функции $f(x)$. Оно является частным случаем (21) при $\bar{\lambda} = \lambda e, \lambda \in R^+$;
3. $\arg \min_{\lambda \rightarrow \infty} \varphi(x, \lambda) \in E$. (25)

В совокупности с предположениями (17) и (22) получаем, что $\varphi(x, \lambda) \in C^2(\square^n)$, т.е. (1),(2),(17),(22) эквивалентна задаче оптимизации (24) дважды непрерывно-дифференцируемой $\varphi(x, \lambda)$.

Функции (23), (24) могут комбинироваться следующим образом

$$\chi(x, \bar{\lambda}, \lambda) = f(x) + \lambda \sum_{i=1}^m f_i^2(x) + \sum_{i=1}^m \lambda_i f_i(x) \xrightarrow{\lambda \rightarrow \infty} \min. \quad (26)$$

Задача (26) может быть решена модифицированным методом множителей Лагранжа [13], который в среднем дает быстрее численное решение исходной задачи по сравнению с методом штрафных функций и одновременно позволяет оценить вектор множителей Лагранжа (28) $\bar{\lambda} = (\lambda_i)_{i \in J_m}$.

Заметим также, что (23)-(26) – задачи оптимизации продолжений $f(x)$ с E , при этом для функций (17) продолжения (23), (26) – непрерывно-дифференцируемые, а (24) – дважды непрерывно-дифференцируемое.

Для иллюстрации здесь и далее выбран класс безусловных булевых задач (the Unconstrained Boolean Problem, UBP):

$$f(x) \rightarrow \min, \quad B_n \quad (27)$$

с булевым множеством $E = B_n$, строгое представление (10) которого имеет вид (14).

Пример 1. Эквивалентные по отношению к (27) задачи (23), (24) имеют, соответственно, вид:

$$\phi(x, \bar{\lambda}) = f(x) + \sum_{i=1}^n \lambda_i (x_i^2 - 1) \rightarrow \min, \quad (28)$$

$$\varphi(x, \lambda) = f(x) + \lambda \sum_{i=1}^n (x_i^2 - 1)^2 \xrightarrow{\lambda \rightarrow \infty} \min. \quad (29)$$

Введем обозначение для штрафного слагаемого

$$\psi(x) = \sum_{i=1}^m f_i^2(x), \quad (30)$$

тогда (24) переписывается так:

$$\varphi(x, \lambda) = f(x) + \lambda \cdot \psi(x) \xrightarrow{\lambda \rightarrow \infty} \min. \quad (31)$$

Пример 2. Для задачи (27), учитывая (14), штрафное слагаемое (30) приобретает вид:

$$\psi(x) = \sum_{i=1}^n (x_i^2 - 1)^2. \quad (32)$$

3.2. Мотивация к использованию выпуклых продолжений

Были предприняты небезуспешные попытки применения метода штрафных функций для решения задач дискретной, в частности булевой, оптимизации [14,15]. Основной проблемой, с которой сталкиваются при этом исследователи, – это большое (порой сравнимое с \mathbb{N}) число локальных минимумов (32) [14]. Теоретически все точки E становятся точками локального минимума $\varphi(x, \lambda)$ при больших λ , так как штрафная часть обнуляется только на E , а значение исходной функции (1) становится несущественным по сравнению с $\lambda \cdot \psi(x)$. Это объясняется невыпуклостью штрафного слагаемого (30) даже для выпуклых строгих представлений множеств.

Так, для B_n^1 (см. (14)) $f_i(x) = x_i^2 - 1$ – выпукла и при этом $f_i(x) < 0$, $x \in (-1,1)$, поэтому $f_i^2(x) = (x_i^2 - 1)^2$ не является выпуклой ($i \in J_n$).

Обобщим это наблюдение в следующем утверждении.

Утверждение. Если E – дискретное множество вида (6) со строгим представлением (10), тогда штрафное слагаемое (30) – невыпукло.

Доказательство от противного. Предположим, (30) – выпуклая, т.е.

$$\forall x^1, x^2 \in \mathbb{R}^n \quad \forall \lambda \in [0,1], \quad \psi(\lambda \cdot x^1 + (1-\lambda) \cdot x^2) \leq \lambda \cdot \psi(x^1) + (1-\lambda) \cdot \psi(x^2). \quad (33)$$

Выберем две различные точки E :

$$x^1, x^2 \in E, \quad x^1 \neq x^2, \quad (34)$$

что возможно в силу (6), и рассмотрим (30) на интервале $(x^1, x^2): x = \lambda \cdot x^1 + (1-\lambda) \cdot x^2, \lambda \in (0,1)$.

Исходя из (30),

– с одной стороны, $\forall x \in \mathbb{R}^n: \psi(x) \geq 0$,

– с другой стороны, поскольку выполнено (10) и, в частности, $\forall j \in J_m \quad f_j(x^1) = f_j(x^2) = 0$, то

$$\psi(x^1) = \psi(x^2) = 0.$$

Если наше предположение верно, в силу (33),

$$\forall \lambda \in (0,1), \quad 0 \leq \psi(x') \leq \lambda \cdot 0 + (1-\lambda) \cdot 0 = 0 \Rightarrow \psi(x') = 0 \Rightarrow \Rightarrow f_j(x') = 0, \quad j \in J_m \Rightarrow x' \in E, \quad (35)$$

так как по условию (10) – строгое функциональное представление E .

Учитывая произвольность выбора $x^1, x^2 \in E$ и объединения (35) с (33), получаем: $\forall x^1, x^2 \in E \quad [x^1, x^2] \in E$, что, очевидно, нарушает (6). Полученное противоречие доказывает утверждение.

Однако известно (см. п.3), что для произвольной функции, в том числе для штрафной функции (31), для заданного λ существует выпуклое продолжение. Как было показано, преимущество использования, вместо $f(x)$, продолжения с E , представляющего собой:

– штрафную функцию, является ею достижение минимума в точке E (см. (25));

– выпуклую функцию – одноэкстремальность и возможность оценить снизу z^* в (1).

Исследуем возможности совместного использования штрафных функций и выпуклых продолжений исходной целевой функции для решения поставленной задачи (1).

Обозначим через $F(x, \bullet)$ – выпуклое продолжение $f(x)$ ($F(x, \bullet)|_E = f(x)$, $F(x, \bullet)$ – выпукла); $\Phi(x, \lambda, \bullet)$ – выпуклое продолжение штрафной функции $\varphi(x, \lambda)$.

Существование выпуклых продолжений функции $f(x)$ было обосновано в п. 3, а поскольку $\Phi(x, \lambda, \bullet)$ – выпукла и одновременно является продолжением $f(x)$, то она также существует.

Заметим, что с дискретных множеств выпуклое продолжение определено далеко неоднозначно. Рассмотрим два способа построения $\Phi(x, \lambda, \bullet)$: а) при помощи выпуклого продолжения (30); б) за счет выпуклого продолжения функции (31) в целом. Сравним два эти подхода.

3.3. Различные способы построения выпуклых продолжений функции $\Phi(x, \lambda)$

Способ 1. Построение $\Phi(x, \lambda, \bullet) = \Phi(x, \lambda)$ за счет выпуклого продолжения штрафной функции

По условию первое слагаемое в (31) – выпукло, а второе – невыпукло по утверждению. Следовательно, для построения $\Phi(x, \lambda)$ достаточно построить выпуклое продолжение $\Psi(x)$ функции $\psi(x)$:

$$\Phi(x, \lambda) = f(x) + \lambda \cdot \Psi(x). \quad (36)$$

Пример 3. Для $E = B_n^1$ построим одно из возможных выпуклых продолжений функции (32), используя функциональное представление (14), в частности, производя для невыпуклых слагаемых замену $x_i^2 = 1$ ($\forall i \in J_n$):

$$\begin{aligned} \psi(x) &= \sum_{i=1}^n (x_i^2 - 1)^2 = \sum_{i=1}^n (x_i^4 - 2 \cdot x_i^2 + 1) = \sum_{i=1}^n x_i^4 - \\ &- 2 \cdot \sum_{i=1}^n x_i^2 + n = \sum_{B_n} x_i^4 - 2 \cdot \sum_{i=1}^n 1 + n = \sum_{i=1}^n x_i^4 - n. \end{aligned} \quad (37)$$

Правая часть (37) – выпуклая, т.е. является выпуклым продолжением функции (32):

$$\Psi(x) = \sum_{i=1}^n x_i^4 - n. \quad (38)$$

Сравним исходную и полученную функции. Функция (32):

– достигает минимума только в точках E :

$$\min_{\square^n} \psi(x) = 0, \quad \text{Arg min}_{\square^n} \psi(x) = E,$$

где $\forall M \subseteq \mathbb{R}^n \quad \text{Arg min}_M f(x) = \left\{ x' \in M : f(x') = \min_M f(x) \right\}$;

– принимает значение n :

в пределах P – в начале координат (при этом O – точка локального максимума);

вне P – в остальных точках множества размещений с повторениями [7, 10] $E' = \bar{E}_3^2(G)$, $G = \{-\sqrt{2}, 0, \sqrt{2}\}^n$:

$$n = \psi(x) = \sum_{i=1}^n (x_i^2 - 1)^2 = \sum_{i=1}^n x_i^4 - 2 \cdot \sum_{i=1}^n x_i^2 + n;$$

$$0 = \sum_{i=1}^n x_i^4 - 2 \cdot \sum_{i=1}^n x_i^2 = \sum_{i=1}^n x_i^2 \cdot (x_i^2 - 2) \Leftrightarrow x \in \{0, \pm\sqrt{2}\}^n.$$

Функция (38):

– достигает минимума в начале координат:

$$\min_{\square^n} \Psi(x) = -n, \quad \text{Arg min}_{\square^n} \Psi(x) = O;$$

– принимает значение 0, помимо E , на целой поверхности четвертого порядка:

$$S = \left\{ x \in \mathbb{R}^n : \sum_{i=1}^n x_i^4 = n \right\} \quad (39)$$

$$(0 = \Psi(x) = \sum_{i=1}^n x_i^4 - n \Leftrightarrow x \in S \supset E).$$

Замечание 2. Итак, функция (39) – выпукла, хотя и достигает своего минимума вне E . С ее помощью выпуклое продолжение (36) штрафной функции (31) приобретает вид:

$$\Phi(x, \lambda) = f(x) + \lambda \cdot \Psi(x) = f(x) + \lambda \cdot \left(\sum_{i=1}^n x_i^4 - n \right). \quad (40)$$

Отсюда видно, что в $\Phi(x, \lambda)$ штраф накладывается только за выход за пределы выпуклого тела

$$C = \text{conv}S, \quad (41)$$

ограниченного поверхностью (39), что является существенной релаксацией по отношению к штрафу за выход с поверхностей (10), заложенный в (30)-(31).

Тот недостаток (38), что $\min_E \Psi(x) \neq \min_{\square^n} \Psi(x)$, а имен-

но $\min_{\square^n} \Psi(x) < \min_E \Psi(x)$, легко снимается переходом

от $\Psi(x)$ к выпуклому продолжению функции $\psi(x)$ вида:

$$\Psi'(x) = \max(\Psi(x), 0) \quad (42)$$

и, соответственно, от (40) к

$$\Phi'(x, \lambda) = f(x) + \lambda \cdot \Psi'(x) = f(x) + \lambda \cdot \max\left(0, \sum_{i=1}^n x_i^4 - n\right)$$

В терминах (41), (42) переписывается в виде:

$$\Psi'(x) = \begin{cases} 0, & x \in C, \\ \sum_{i=1}^n x_i^4 - n, & x \notin C. \end{cases}$$

Соответственно, мы имеем два случая: если решение безусловной задачи:

1) внутренняя точка C , в том числе P :

$$x^{**} = \arg \min_{\square^n} f(x) \in C \Rightarrow \max\left(0, \sum_{i=1}^n x_i^{**4} - n\right) = 0, \quad (43)$$

то задача минимизации $\Phi'(x, \lambda)$ фактически сводится к поиску x^{**} , т.е. к решению релаксационной по отношению к (1)-(2) задачи на $C \supset E$ вместо E ;

2) если x^{**} – внешняя точка C , то $\sum_{i=1}^n x_i^{**4} - n > 0$ и в

результате применения метода штрафных функций для $\Phi'(x, \lambda)$ мы получим $x' \in S$, т.е. опять же решим лишь релаксационную по отношению к (1)-(2) задачу.

Как видно, использование выпуклого продолжения для штрафного слагаемого нецелесообразно, поскольку приводит к оптимизации функции, достигающей минимума вне E , и снимает главное требование, заложенное в нем, – принадлежность E .

Способ 2. Построение $\Phi(x, \lambda, \mu(\lambda))$ за счет выпуклого продолжения исходной целевой функции.

Естественным образом мы пришли к идее использования выпуклого продолжения функции (31) за счет выпуклого продолжения функции $f(x)$ при неизменном штрафном слагаемом:

$$\varphi(x, \lambda) \rightarrow \Phi(x, \lambda, \mu(\lambda)) = F(x, \mu(\lambda)) + \lambda \cdot \psi(x),$$

$$F(x, \mu(\lambda)) = f(x) + \mu(\lambda) \cdot \gamma(x): \quad (44)$$

а) $F(x, \mu(\lambda))$ – выпуклое продолжение $f(x)$, обеспечивающее выпуклость $\Phi(x, \lambda, \mu(\lambda))$ для заданного λ ,

в) $\gamma(x)$ – корректирующая функция. Так, для рассматриваемых нами полиэдрально-сферических множеств корректирующая функция имеет вид:

$$\gamma(x) = \mu(\lambda) \cdot \left((x-a)^2 - r^2 \right), \quad (45)$$

соответственно,

$$\Phi(x, \lambda, \mu(\lambda)) = f(x) + \mu(\lambda) \cdot \left((x-a)^2 - r^2 \right) + \lambda \cdot \psi(x) \quad (46)$$

выпукла.

Перепишем (46) в виде:

$$\Phi(x, \lambda, \mu(\lambda)) = \tilde{f}(x, \lambda) + \mu(\lambda) \cdot \left((x-a)^2 - r^2 \right), \quad (47)$$

где
$$\tilde{f}(x, \lambda) = f(x) + \lambda \cdot \sum_{i=1}^m f_i^2(x). \quad (48)$$

Функция (49) имеет свойства, аналогичные $f(x)$ (кроме выпуклости):

– если (17) выполнено, то, в силу (22),

$$\forall |\lambda| < \infty: \tilde{f}(x, \lambda) \in C^2(\square^n); \quad (49)$$

– если выполнено (20) и

$$\frac{\partial f_j(x)}{\partial x_i} < \infty, \quad i \in J_n, \quad j \in J_m,$$

то
$$\frac{\partial^2 \tilde{f}(x, \lambda)}{\partial x_i \cdot \partial x_j} < \infty, \quad i, j \in J_n. \quad (50)$$

Соответственно, согласно [18], при выполнении (49) существует выпуклое продолжение $\tilde{f}(x, \lambda)$ в форме (48) на любой компакт $E' \supset E$. Если к тому же выполнено (50), такое продолжение существует с E в \square^n . Более того, существует целое семейство таких продолжений:

$$\forall \lambda^* \in \square \exists M(\lambda^*) = g \left(\frac{\partial^2 \tilde{f}(x, \lambda^*)}{\partial x_i \cdot \partial x_j} \right)_{i, j \in J_n} \geq 0:$$

$\forall \mu(\lambda) \geq M(\lambda^*)$ функция (47) – выпукла.

В то же время, поскольку $\mu(\lambda) \geq 0$, то в (47)-(48)

только составляющая $\lambda \cdot \sum_{i=1}^m f_i^2(x)$ невыпукла, откуда

$$\forall \lambda \leq \lambda^* \Phi(x, \lambda, M(\lambda^*)) = \tilde{f}(x, \lambda) + M(\lambda^*) \cdot \left((x-a)^2 - r^2 \right)$$

будет выпуклой.

Замечание 3. Для определения $M(\lambda^*)$ могут быть применены оценки, приведенные в [18]. Также могут быть использованы свойства конкретной функции (30).

Так, для B_n' (45) корректирующая функция имеет вид

$$\gamma(x) = \sum_{i=1}^n x_i^2 - n, \text{ соответственно (47) может быть пре-}$$

образована так: $\Phi(x, \lambda^*, M(\lambda^*)) =$

$$\begin{aligned} &= f(x) + M(\lambda^*) \cdot \left(\sum_{i=1}^n x_i^2 - n \right) + \lambda^* \cdot \sum_{i=1}^n (x_i^2 - 1)^2 = \\ &= f(x) + M(\lambda^*) \cdot \left(\sum_{i=1}^n x_i^2 - n \right) + \lambda^* \cdot \left(\sum_{i=1}^n x_i^4 - 2 \sum_{i=1}^n x_i^2 + n \right) = \\ &= f(x) + \lambda^* \cdot \sum_{i=1}^n x_i^4 + \left(M(\lambda^*) - 2\lambda^* \right) \sum_{i=1}^n x_i^2 + \\ &+ \left(M(\lambda^*) - 2\lambda^* \right) n. \end{aligned}$$

Отсюда видно, что для UBP (27) достаточно выбрать

$$M(\lambda^*) = 2 \cdot \lambda^*.$$

Главное отличие в сформированных выпуклых продолжениях $\varphi(x, \lambda)$ и, соответственно, $f(x)$ состоит в том, что первое (36) является выпуклым продолжением штрафной функции $f(x)$, в то время как второе

$$\Phi(x, \lambda, \mu(\lambda^*)) = f(x) + \mu(\lambda^*) \cdot ((x-a)^2 - r^2) + \lambda \cdot \psi(x)$$

– это штрафная функция выпуклого продолжения $f(x)$.

3.4. Описание метода

Зафиксируем произвольное, сколь угодно большое $\lambda^* > 0$ и определим $M(\lambda^*)$ (см. замечание 3). Исходная задача будет эквивалентна задаче оптимизации штрафной функции:

$$\Phi(x, \lambda, M(\lambda^*)) = f(x) + M(\lambda^*) \cdot ((x-a)^2 - r^2) + \lambda \cdot \sum_{i=1}^m f_i^2(x) \rightarrow \min_{\lambda \rightarrow \infty} \quad (51)$$

Применим метод штрафных функций к $\Phi(x, \lambda, M(\lambda^*))$ для последовательности штрафных параметров $\lambda \in \Lambda$:

$$\Lambda = \{\lambda^k\}_k \geq 0: \lambda^0 = 0, \lambda^{k-1} < \lambda^k, k \in \mathbb{N}.$$

Обозначим z^l, z^u нижнюю и верхнюю оценки z^* (см.(1)), соответственно.

Для получения z^l используем тот факт, что $\forall \lambda^k \leq \lambda^*$ $\Phi(x, \lambda^k, M(\lambda^*))$ – выпукла и решение x^k задачи (51) дает нижнюю оценку целевой функции:

$$z^l \geq \max_{\lambda^k \leq \lambda^*} \Phi(x^k, \lambda^k, M(\lambda^*)). \quad (52)$$

Обозначим через y^k – проекцию x^k на E ($y^k = \text{Pr}_E x^k$) или результат комбинаторного округления x_k до элемента E ($y^k = [x^k]_E$) [4], тогда верхняя оценка:

$$z^u \leq \min_{\lambda^k \in \Lambda} f(y^k). \quad (53)$$

Замечание 4. Для полиэдрально-сферических множеств, рассматриваемых нами, квадратичная задача поиска проекции эквивалентна линейной задаче, которая и для многих ЕКМ решается в явном виде [7] или, по крайней мере, эффективно (такие множества называются хорошо описанными [28]).

Пример 4. Для B_n^1 все множество проекций на него:

$$Y^k = \{y^k = \text{Pr}_E x^k\} = \begin{cases} -1, & \text{если } x_i^k < 0, \\ \pm 1, & \text{если } x_i^k = 0, (i \in J_n), \\ 1, & \text{если } x_i^k > 0, \end{cases}$$

В зависимости от того, удовлетворяет ли нас точность достигнутого решения, процесс можно завершить либо повторить для большего λ^* .

Замечание 5. Также можно сформировать серию траекторий допустимых решений вспомогательной задачи (51) для $\lambda^* \in \Lambda^*$:

$$\Lambda^* = \{\lambda^{*j}\}_j \geq 0: \lambda^{*0} = 0, \lambda^{*(j-1)} < \lambda^{*j}, j \in \mathbb{N}.$$

В обозначениях

$$x^{jk} = \arg \min \Phi(x, \lambda^k, M(\lambda^{*j})), y^{jk} = \text{Pr}_E x^{jk}, j, k,$$

оценки (52), (53) принимают вид:

$$z^l \geq \max_{\substack{\lambda^k \leq \lambda^{*j} \\ \lambda^{*j} \in \Lambda^*, \lambda^k \in \Lambda}} \Phi(x^{jk}, \lambda^k, M(\lambda^{*j})); \quad (54)$$

$$z^u \leq \min_{\lambda^{*j} \in \Lambda^*, \lambda^k \in \Lambda} f(y^{jk}). \quad (55)$$

В результате будет сформирован набор траекторий:

$$X^j = \{x^{jk}\}_k, j, \text{ с началом в точках } x^{j0} = \arg \min \Phi(x, 0, M(\lambda^{*j})) \text{ и концами в точках } E.$$

Начальные точки $\{x^{j0}\}_j$ также формируют траекто-

рию, которая стремится к центру сферы (4): $x^{j0} \xrightarrow{j \rightarrow \infty} a$

за счет увеличения параметра $M(\lambda^{*j})$ в $\Phi(x, \lambda, M(\lambda^{*j})) = f(x) + M(\lambda^{*j}) \cdot ((x-a)^2 - r^2)$ с увеличением j ($M(\lambda^{*j}) \xrightarrow{j \rightarrow \infty} \infty$).

Поскольку мы рассматриваем полиэдрально-сферические множества, все точки которых равноудалены от центра a описанной сферы, на этапе проектирования будет получено все множество E и таким образом теоретически задача (1) будет решена точно.

Практически же процесс решения прерывается по достижению заданной точности ε по текущим, ниж-

ней и верхней, оценкам: $\frac{z^u - z^l}{z^l} < \varepsilon$.

Заметим также, что нижняя оценка может быть существенно улучшена при помощи решения релаксационных задач (см. п.3.1) на сфере и многограннике:

$$z^u \geq \max \left\{ \min_P f(x), \min_{S_r(a)} f(x) \right\}.$$

Эти вспомогательные задачи эффективно решаются, в частности, на хорошо описанных комбинаторных множествах и для таких функций как квадратичные [28].

Выводы

Представлен новый подход к решению задач оптимизации для полиэдрально-сферических комбинаторных множеств, основанный на применении, наряду с методом штрафных функций, функциональных представлений евклидовых комбинаторных множеств и выпуклых продолжений функций с этих множеств в \mathbb{R}^n .

Данный метод может быть обобщен на произвольные хорошо описанные вершинно расположенные евклидовые комбинаторные множества при наличии их строгих функциональных представлений.

Непосредственное обобщение метода на не вершинно расположенные множества невозможно, поскольку основано на построении выпуклых продолжений функций, что возможно для произвольной функции только для вершинно расположенных множеств. Однако после предварительных преобразований, сводящих исходную задачу к задаче на вершинно расположенном множестве в пространстве большей размерности (см., например, в [16] схему сведения произвольной целочисленной и дискретной задачи с ограничениями, переменными к булевой задаче), данный метод применим.

Таким образом, актуальным является развитие теории выпуклых продолжений, решение линейных задач, а также поиск функциональных представлений различных вершинно расположенных комбинаторных множеств, таких как общее множество перестановок, вершинно расположенные классы размещений и сочетаний и их композиционные образы.

Литература: 1. Пападимитриу Х., Стайглиц К. Комбинаторная оптимизация. Алгоритмы и сложность. М.: Мир, 1984. 512 с. 2. *Approximation and Complexity in Numerical Optimization: Continuous and Discrete Problems* / edited by P.M. Pardalos. Boston: Kluwer Academic Publishers, 2000. 581 p. 3. Kochenberger G., Hao J.-K., Glover F., Lewis M., Lu Z., Wang H., and Wang Y. The unconstrained binary quadratic programming problem: a survey // *Journal of Combinatorial Optimization*. 2014. №1. P. 58-81. 4. Емеличев В. А., Ковалев М. М., Кравцов М. К. Многогранники, графы, оптимизация (комбинаторная теория многогранников). М.: Наука, 1981. 344 с. 5. Сергиенко И.В., Гуляницкий Л.Ф., Сиренко С.И. Классификация прикладных методов комбинаторной оптимизации // *Кибернетика и системный анализ*. 2009. № 5. С. 71-83. 6. Стоян Ю.Г., Яковлев С.В. Математические модели и оптимизационные методы геометрического проектирования. К.: Наук. думка, 1986. 268 с. 7. Стоян Ю. Г., Ємець О. О. Теорія і методи евклідової комбінаторної оптимізації. К.: Ін-т системн. дослідж. освіти, 1993. 188 с. 8. Стоян Ю.Г. Некоторые свойства специальных комбинаторных множеств: Препринт 85 АН УССР. Х.: Институт проблем машиностроения, 1980. 22 с. 9. Стоян Ю.Г., Яковлев С.В., Паршин О.В. Квадратичная оптимизация на комбинаторных множествах в \mathbb{R}^n // *Кибернетика и системный анализ*, 1991. №4. С. 97-104. 10. Яковлев С.В., Гребенник И.В. О некоторых классах задач оптимизации на множестве размещений и их свойствах // *Изв. вузов. Математика*, 1991. №11. С. 74-86. 11. Пичугина О.С., Яковлев С.В. Полиэдрально-сферический подход к решению некоторых классов комбинаторных задач // *Праці VI Міжнародної школи-семінару «Теорія прийняття рішень»*. Ужгород: УжНУ, 2012. С. 152-

153. 12. Ємець О.О., Ємець С.М. Відсікання в лінійних частково комбінаторних задачах евклідової комбінаторної оптимізації // *Доп. НАН України*. 2000. № 9. С. 105-109. 13. Bertsekas D.P. *Nonlinear Programming*. Belmont: Athena Scientific. 1995. 378 p. 14. Murray W., Ng K.-M. An algorithm for nonlinear optimization problems with binary variables // *Computational Optimization and Applications*, 2010. V. 47. P. 257-288. 15. Chen J.-S., Li J.-F., Wul J. A continuation approach for solving binary quadratic program based on a class of NCP-functions // *Applied Mathematics and Computation*, 2012. V. 219. P. 3975-3992. 16. Писарук Н.Н. Модели и методы смешанно-целочисленного программирования. Минск: Изд-во БГУ, 2008. 250 с. 17. Яковлев С. В. Теория выпуклых продолжений функций на вершинах выпуклых многогранников // *Журнал вычислительной математики и математической физики*. 1994. Т. 34, № 7. С. 1112-1119. 18. Стоян Ю. Г., Яковлев С. В. Построение выпуклых и вогнутых функций на перестановочном многограннике // *ДАН УССР, СЕР. А*. 1988. № 5. С. 66-68. 19. Стоян Ю.Г., Яковлев С.В., Ємець О.А., Валуйская О.А. Построение выпуклых продолжений для функций, заданных на гиперсфере // *Кибернетика и системный анализ*, 1998. № 2. С. 27-37. 20. Postnikov A. Permutohedra, Associahedra, and Beyond // *IMRN: International Mathematics Research Notices*, 2009. №6. P. 1026-1106. 21. Гребенник И. В., Литвиненко А. С. Генерация комбинаторных множеств с заданными свойствами // *Кибернетика и системный анализ*. 2012. № 6. С. 96-105. 22. Krislock N., Malick J., Roupin F. Computational results of a semidefinite branch-and-bound algorithm for k-cluster // *Computers & Operations Research*, 2016. №66. P. 153-159. 23. Billionnet A., Elloumi S., Plateau M.-C. Improving the performance of standard solvers for quadratic 0-1 programs by a tight convex reformulation: The QCR method // *Discrete Applied Mathematics*, 2009. №157. P. 1185-1197. 24. Brualdi R. A. *Combinatorial matrix classes*. Cambridge: Cambridge University Press, 2006. 544 p. 25. Burkard R. E., Pardalos P. M., Du D.-Z., Graham R. L. Quadratic Assignment Problems // in *Handbook of Combinatorial Optimization*. New York: Springer, 2013. P. 2741-2814. 26. Hillier F. S., Appa G., Pitsoulis L., Williams H. P., Pardalos P. M., Prokopyev O. A., Busygin S. Continuous Approaches for Solving Discrete Optimization Problems // in *Handbook on Modelling for Discrete Optimization*. New York: Springer, 2006. P. 1-39. 27. Punnen A. P., Sripratak P., Karapetyan D. The bipartite unconstrained 0-1 quadratic programming problem: Polynomially solvable cases // *Discrete Applied Mathematics*, 2015. V. 193. P. 1-10. 28. Bernstein Y., Lee J., Onn S., Weismantel R. Parametric nonlinear discrete optimization over well-described sets and matroid intersections // *Mathematical Programming*, 2010. V. 124, №1/2. P. 233-253. 29. Dahl J. *Convex problems in signal processing and communications* // Ph.D. thesis, Aalborg University. 2003. P.100.

Поступила в редколлегию 18.02.2016

Рецензент: д-р физ.-мат. наук, проф. Колечкина Л.Н.

Пичугина Оксана Сергеевна, канд. физ.-мат. наук, докторант кафедры прикладной математики ХНУРЕ. Научные интересы: полиэдральная комбинаторика, евклидовая комбинаторная, нелинейная, параметрическая оптимизация, теория графов. Адрес: Украина, 61166, Харьков, пр. Науки, 14, тел. (099)9598965.

Яковлев Сергей Всеволодович, д-р физ.-мат. наук, профессор кафедры информационных технологий и защиты информации Харьковского национального университета внутренних дел. Научные интересы: нелинейная оптимизация, евклидовая комбинаторная оптимизация, задачи упаковки и раскроя. Адрес: Украина, 61080, Харьков, пр. 50-летия СССР, 27, тел. (050)8046392.

ДИНАМІЧНА МОДЕЛЬ ОПТИМІЗАЦІЇ ТРАНСПОРТНИХ ВИТРАТ З УРАХУВАННЯМ НЕВИЗНАЧЕНОСТІ ТЕРМІНІВ ДОСТАВКИ

НОВОЖИЛОВА М.В., ЛИТВИНЕНКО Є.М.

Розглядається транспортно-логістична система обслуговування клієнтів з організацією доставки та вивезення інструментів. Пропонується динамічна модель виконання портфеля замовлень на визначеному горизонті планування з мінімізацією транспортних витрат.

1. Вступ

Перевезення продукції або обладнання автомобільним транспортом є невід'ємною складовою функціонування будь-якого підприємства або організації. В умовах конкуренції та жорсткої обмеженості ресурсів суб'єкти господарювання приділяють значну увагу мінімізації транспортних витрат.

Особливо актуальною задачею мінімізації транспортних витрат є для транспортно-логістичних систем (ТЛС) підприємств та організацій [1], які залежать від щоденного використання транспортних засобів, наприклад:

– надають клієнтам, на їх замовлення, в оренду інструменти, машини та механізми або зразки інструментів для тимчасового користування та випробування;

– організовують тестування та обслуговування клієнтського обладнання у своїх лабораторіях тощо.

Важливою умовою функціонування таких транспортно-логістичних систем є організація доставки, на замовлення клієнтів, обладнання (інструментів певних типів) з їх наступним поверненням через деякий проміжок часу. Крім того, для таких систем характерна можливість відтермінування у часі виконання замовлень, узгоджена з клієнтами.

Складність, багатовимірність і, в загальному випадку, багатокритеріальність задач обумовлює необхідність моделювання ТЛС із урахуванням можливості зміни часу виконання замовлень з метою:

– зменшення витрат на утримання транспортних засобів шляхом мінімізації їх кількості;

– зменшення витрат на слідування до клієнта шляхом оптимізації транспортних маршрутів;

– зменшення холостого пробігу транспортних засобів.

2. Огляд попередніх досліджень

Починаючи з відомої публікації Г. Данцига [2], яка була першою в даній предметній галузі, цей клас задач викликає незмінний інтерес дослідників. Задачі оптимізації структури та складу ТЛС розглядалися у чис-

ленних наукових публікаціях вітчизняних та закордонних авторів.

Виділимо роботи [3, 4], які присвячені задачам оперативного планування перевезень та розподілу вантажів. В дослідженні [5] запропоновано багатокритеріальну задачу підвищення ефективності управління перевезенням вантажів у динамічній транспортній мережі за рахунок складання попереднього плану перевезень.

В роботах [6-10] розглянуті задачі маршрутизації транспортних засобів. В [7] досліджена транспортна модель з обмеженою кількістю транспортних засобів (ТЗ), в публікації [9] розглянуто наближений метод розв'язання кластерної задачі маршрутизації.

Слід зазначити, що задачі, які передбачають не тільки доставку, але й повернення вантажів, є вкрай малодослідженими. Крім того, в більшості досліджень розглядаються задачі оперативного планування [11, 12], а задачам з необхідністю використання горизонту планування приділяється недостатня увага.

3. Опис транспортно-логістичної системи

Метою дослідження є побудова динамічної моделі задачі мінімізації витрат в ТЛС з урахуванням можливості відтермінування у часі виконання замовлень.

Розглянемо ТЛС як цілеспрямовану систему Λ , що містить множину різнорідних елементів M , на якій реалізовано множину зв'язків R , що упорядковують елементи в структуру

$$\Lambda = \langle (M \times R) \times P \rangle, \quad (1)$$

де P – множина властивостей структури $(M \times R)$, що дають змогу досягти заданої цілі [13].

В системі, що розглядається, множина елементів M містить склад C_0 , множину клієнтів $C = (C_1, C_2, \dots, C_N)$ та множину транспортних засобів $V = (V_1, V_2, \dots, V_J)$ (рис. 1).

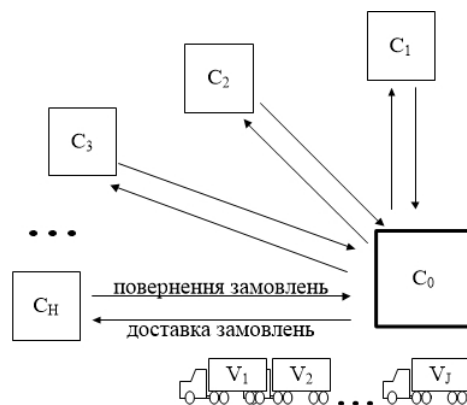


Рис. 1. Елементи ТЛС

Позначимо координати розміщення C_0 через (x_0, y_0) . Склад C_0 містить упорядковану множину типів інструментів $I = (I_1, I_2, \dots, I_N)$, які можуть бути надані клієнтам на їх замовлення.

Кожен з типів інструментів визначається параметрами:

$$I_i = (q_i, w_i), i = 1, \dots, I,$$

де q_i – доступна кількість інструментів i -го типу; w_i – вага одиниці інструмента i -го типу.

Для використання у подальшому введемо

$$Q = \sum_{i=1}^I q_i, Q^i = \sum_{n=1}^{i-1} q_n.$$

Позначимо через (x_h, y_h) координати розміщення елемента $C_h, h=1, \dots, N$ множини клієнтів C .

Визначення 1. Горизонтом планування $[1, T]$ будемо називати інтервал часу, на якому планується виконання замовлення клієнтів.

Врамак даної роботи вважатимемо елементи множини $V = (V_1, V_2, \dots, V_J)$ однотипними та визначатимемо як

$$V_j = (AW_j, AD_j, AC_j), j = 1, \dots, J, \quad (2)$$

де AW_j – вантажопідйомність; AD_j – максимальний добовий пробіг; AC_j – вартість використання V_j за період $[1, T]$.

Транспортні засоби виділяються на весь період горизонту планування. Використання ТЗ один день означає його оренду на весь період функціонування системи.

Транспортні перевезення здійснюються згідно з транспортною мережею, яка є повним графом $G = (C, E)$ з множиною вершин C та множиною ребер E . Вершинам графа відповідають склад та клієнти $C = \{C_0, C_h\}$, а ребрам $e_{ug} \in E$ – ділянки шляху, які їх з'єднують (рис. 2).



Рис. 2. Граф транспортної мережі

Ваги s_{ug} ребер e_{ug} обчислюються таким чином:

$$s_{ug} = \sqrt{(x_g - x_u)^2 + (y_g - y_u)^2}, \quad (3)$$

$u, g = 0, \dots, N, u \neq g$.

Транспортний засіб може здійснювати декілька рейсів протягом одного дня.

Визначення 2. Рейсом 1 транспортного засобу V_j будемо називати цикл на графі G вигляду

$$l = (C_0, \dots, C_h, C_{h+1}, \dots, C_0).$$

Рейс є припустимим, якщо виконані замовлення усіх поточних клієнтів та загальна вага інструментів, що перевозяться, не перевищує вантажопідйомність AW_j .

Визначення 3. Маршрутом Ψ_{jt} назвемо сукупність рейсів V_j , які виконуються протягом одного дня t .

Маршрут Ψ_{jt} є припустимим, якщо він складається з припустимих рейсів та їх загальна довжина не перевищує максимальний добовий пробіг транспортного засобу AD_j .

Замовлення інструментів клієнтами визначаються як:

$$R_k = (i_k, h_k, t_{\min k}, t_{\max k}, q_k, d_k), k = 1, \dots, K, \quad (4)$$

де k – номер замовлення; $i_k = 1, \dots, N$ – тип замовленого інструмента; $h_k = 1, \dots, N$ – номер клієнта; $[t_{\min k} \dots t_{\max k}]$ – період виконання замовлення; q_k – кількість замовлених інструментів; d_k – тривалість виконання замовлення.

Замовлення $R = (R_1, R_2, \dots, R_K)$ утворюють множину зв'язків ТЛС (1).

Таким чином, задача планування роботи ТЛС на період $[1, T]$ включає виконання K замовлень на доставку клієнтам інструментів зі складу та їх повернення на склад.

4. Властивості постановки задачі

Визначимо властивості постановки задачі, що розглядається.

Властивість 1. Кожний клієнт C_h генерує одне або декілька замовлень R_k виду (4). Замовлення містить інструменти одного типу. У разі необхідності замовлення декількох інструментів клієнт генерує відповідну кількість замовлень.

Властивість 2. Момент t_k початку виконання певного замовлення R_k належить інтервалу $[t_{\min k}, t_{\max k} - d_k + 1]$.

Згідно з властивістю 2 існує $(t_{\max k} - d_k - t_{\min k} + 1)$ варіантів виконання замовлення R_k (рис. 3).

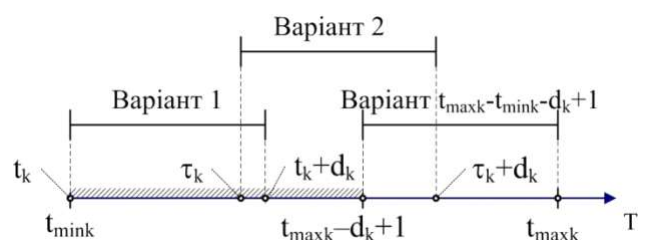


Рис. 3. Приклади варіантів виконання замовлення R_k

Властивість 3. Повернення інструментів типу i_k від клієнта C_h на склад відбувається у перший день після завершення виконання замовлення R_k . Це день $(t_k + d_k)$. Інструмент стає доступним для іншого замовлення наступного дня $(t_k + d_k + 1)$.

Зважаючи на введені визначення та виділені властивості змістовної постановки задачі, побудуємо загальну динамічну модель оптимізації транспортних витрат.

5. Динамічна модель оптимізації транспортних витрат

Характеристики та властивості ТЛС, наведені вище, є фундаментом постановки задачі мінімізації транспортних витрат ТЛС.

Припустимим (але, очевидно, неоптимальним) є рішення, при якому кожний рейс засобу V_j виконує тільки одне замовлення (відвідує одного клієнта), а кожний транспортний засіб виконує тільки один рейс в день. Одне замовлення, у разі перевищення ним вантажопідйомності ТЗ, може виконуватися декількома ТЗ, кількість яких можна визначити як

$$J_k = \text{round}\left(\frac{q_k^i \times w_i}{AW_j}\right),$$

де round – округлення вгору до найближчого цілого числа.

Тоді верхня оцінка J_{\max} кількості J , необхідних ТЗ, дорівнює

$$J_{\max} = \frac{\sum_{k=1}^K J_k \times (s_{0h_k} + s_{hk0})}{AD_j}.$$

Ендогенні змінні моделі. Введемо булеві змінні:

$$\beta_{jt} = \begin{cases} 1, \text{ якщо } V_j \text{ використовується у день } t, \\ 0, \text{ інакше,} \end{cases}$$

$$j = 1, \dots, J_{\max},$$

$$\mu_{jt} = \begin{cases} 1, \text{ якщо рейс } l \text{ використовується у день } t, \\ 0, \text{ інакше,} \end{cases}$$

$$l = 1, \dots, L, t = 1, \dots, T,$$

$$x_{ug}^l = \begin{cases} 1, \text{ якщо ребро } e_{ug} \text{ входить у рейс } l, \\ 0, \text{ інакше,} \end{cases}$$

$$l = 1, \dots, L, u, g = 0, \dots, H, u \neq g.$$

Використовуючи підхід [14, 15], введемо у розгляд простір ресурсів TOQ з прямокутною системою координат, в якій горизонтальна вісь OT відповідає горизонту планування $[1, T]$, а вертикальна вісь OQ відбиває інформацію про кількість інструментів i -го типу $i=1, \dots, I$ (рис. 4).

Тоді замовлення R_k є прямокутником розмірів (d_k, q_k^i) у відповідній смузі ресурсів простору TOQ , (t_k, r_k^i) – координати нижнього лівого кута прямокутника R_k , котрі задають розміщення відповідного замовлення у просторі ресурсів (див. рис. 4).

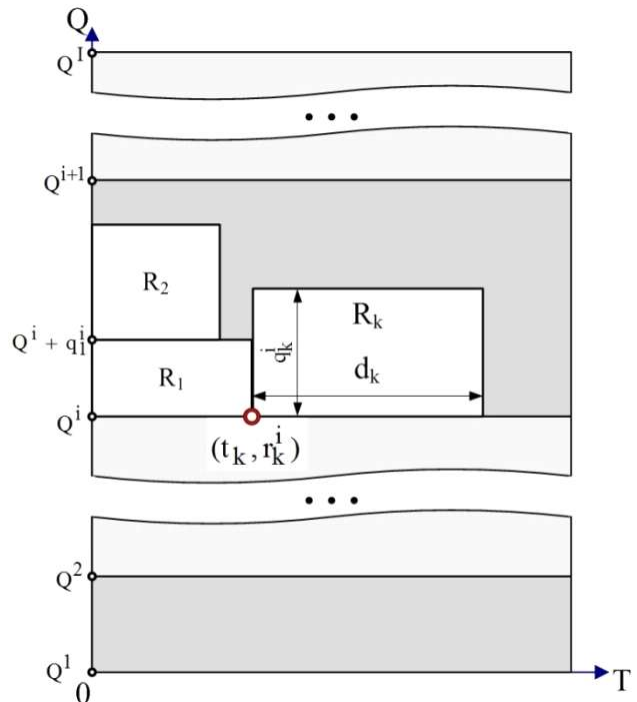


Рис. 4. Розміщення замовлень у просторі ресурсів

Нехай також змінні

$$a_i = \max_{t=1, \dots, T} (r_k^i + q_k^i), \quad i = 1, \dots, N,$$

які визначають максимальну кількість інструментів i -го типу, що одночасно використовуються.

Відмітимо, що ендогенні змінні t_k, r_k^i, a_i можуть сприйматися як неперервні величини, задані у визначених діапазонах.

Цільовий функціонал задачі мінімізації транспортних витрат ТЛС є векторним:

$$F(F_1, F_2, F_3), \quad (5)$$

де функція $F_1(\beta_{jt})$ має вигляд:

$$F_1(\beta_{jt}) = \max_{t=1, \dots, T} \sum_{j=1}^{J_{\max}} \beta_{jt},$$

визначає отримання оптимальної множини транспортних засобів та збігається до мінімуму:

$$F_1(\beta_{jt}) \rightarrow \min.$$

Функціонал $F_2 = \{f_{21}, f_{22}, \dots, f_{2T}\}$, де частковий критерій f_{2t} має вигляд

$$f_{2t} = \sum_{l=1}^L \left(\mu_{lt} \times \left(\sum_{u=0}^H \sum_{g=0}^H s_{ug} \times x_{ug}^l \right) \right) \rightarrow \min,$$

відповідає побудові припустимих щоденних маршрутів мінімальної довжини, які забезпечують виконання передбачених замовлень.

Нарешті, функція F_3 навантаження на транспортні засоби має вигляд

$$F_3 = \sum_{i=1}^N a_i \times w_i \rightarrow \min.$$

Основні обмеження моделі. Загальна кількість доступних для використання інструментів обмежена:

$$a_i \leq q_i. \quad (6)$$

Для виконання кожного замовлення мають бути наявними відповідні інструменти:

$$\begin{cases} r_k^i \geq Q^i \\ -r_k^i \geq -a_i, k = 1, \dots, K. \end{cases} \quad (7)$$

Термін виконання замовлення обмежений:

$$\begin{cases} t_k \geq t_{\min k} \\ -t_k \geq -t_{\max k} + d_k - 1, k = 1, \dots, K. \end{cases} \quad (8)$$

Жодна пара замовлень не може одночасно використовувати один і той же інструмент:

$$\begin{cases} r_m^i - r_k^i \geq q_k^i \\ r_k^i - r_m^i \geq q_m^i \\ t_k - t_m \geq d_m \\ t_m - t_k \geq d_k \end{cases}, k, m = 1, \dots, K. \quad (9)$$

Максимальна довжина маршруту транспортного засобу обмежена:

$$\mu_{lt} \times \sum_{l=1}^L \sum_{u=0}^H \sum_{g=0}^H s_{ug} \times x_{ug}^l \leq AD_j, \forall t = 1, \dots, T, \quad (10)$$

$$\forall j = 1, \dots, J_{\max}.$$

У кожному рейсі кількість ребер, що входить у вершину, дорівнює кількості ребер, які виходять з неї:

$$\sum_{g=0}^H x_{ug}^l = \sum_{g=0}^H x_{gu}^l, u = 0, \dots, H, l = 1, \dots, K. \quad (11)$$

В кожний рейс завжди входить склад:

$$\sum_{g=0}^H x_{og}^l = 1, u = 0, \dots, H, l = 1, \dots, K. \quad (12)$$

Задача (5)-(12) є багатокритеріальною та NP-складною, для розв'язання якої пропонується розбити її на дві взаємозалежні задачі.

Задача 1. Пошук раціонального $(t_1, r_1^i, t_2, r_2^j, \dots, t_K, r_K^n)$ розміщення замовлень на горизонті планування $[1, T]$, або можливо у визначений день t .

Задача 2. Пошук маршруту виконання замовлень для кожного часового такту $t = 1, \dots, T$.

Задача 1 визначає на горизонті планування $[1, T]$ час виконання операцій доставки та повернення інструментів на замовлення клієнтів.

Розв'язання задачі 2 дозволяє визначити кількість транспортних засобів, необхідних для виконання усіх замовлень, та розрахувати їх оптимальні маршрути. У разі потреби відбувається виклик задачі 1 для оптимізації розміщення замовлень на деякому періоді горизонту $[1, T]$ (рис. 5).

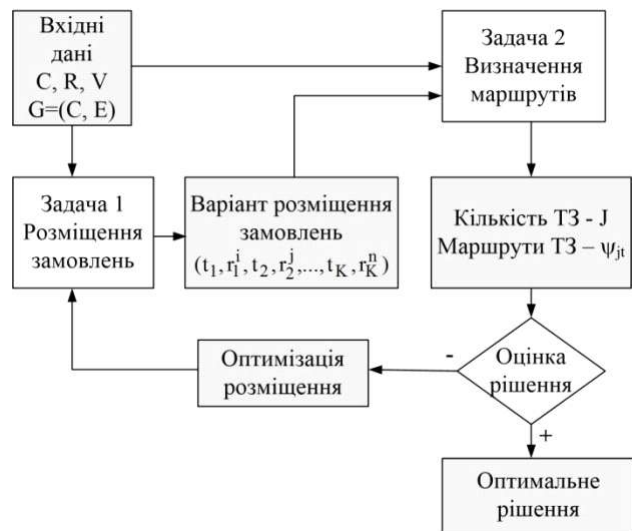


Рис. 5. Розв'язання загальної задачі мінімізації транспортних витрат

6. Висновки

1. Запропоновано теоретико-множинний опис ТЛС, функціонування якої розглядається на горизонті планування.
2. Розроблено багатокритеріальну математичну модель оптимізації транспортних витрат підприємства за рахунок зменшення витрат на оренду ТЗ та оптимізації щоденних транспортних маршрутів. Представлена математична модель дозволяє виконувати стратегічне планування функціонування ТЛС та враховує можливість відтермінування початку виконання замовлень.
3. Запропоноване розбиття оптимізаційної задачі на дві взаємозалежні: задачу розміщення та задачу маршрутизації.

Подальші дослідження спрямовані на побудову та реалізацію методу розв'язання поставленої задачі оптимізації.

Література: 1. *Миротин Л.Б., Гудков В.А., Зырянов В.В.* Управление грузовыми потоками в транспортно-логистических системах / Под ред. Л.Б. Миротина. М.: Горячая линия-Телеком. 2010. 704 с. 2. *Dantzig G.B., Ramser J.H.* The Truck Dispatching Problem // *Management Science*. 1959. Vol. 6, No. 1. P. 80-91. 3. *Гришко С.В., Гуца О.М., Сухомлинов А.И.* Моделивання виробничої логістики в умовах перебудови підприємства // *Радіоелектроніка та інформатика*. 2013. № 2. С. 31-34. 4. *Гвоздинський А.М., Обізна М.Ю.* Дослідження інтелектуальних методів розв'язання оптимізаційних задач транспортного типу // *Радіоелектроніка та інформатика*. 2013. № 4. С. 35-39. 5. *Задоров В.Б., Федусенко О.В., Федусенко А.О.* Застосування методів багатокритеріальної оптимізації до планування вантажних перевезень // *Управління розвитком складних систем*. 2010. №2. С. 27-31. 6. *Меламед И.И., Сергеев С.И., Сигал И.Х.* Задача коммивояжера. Вопросы теории // *Автоматика и телемеханика*. 1989. № 9. С. 3-33. 7. *Hajghasem M., Shojaie A.A.* Optimal routing in supply chain aimed at minimizing vehicle cost and supply // *Procedia Economics and Finance*. 2016. Vol. 36. P. 353-362. 8. *Szucs G.* Decision support for route search and optimum finding in transport networks under uncertainty // *Journal of Applied Research and Technology*. 2015. Vol. 13. P. 125-134. 9. *Кузнецов К.А., Громов В.А.* Подход к решению задачи оптимизации структуры дистрибуторской компании // *Вісник АМСУ. Серія: «Технічні науки»*. 2014. № 1 (51). С. 94-103. 10. *Lang Z., Yao E., Hu W., Pan Z.* A vehicle routing problem solution considering alternative stop points // *Procedia-Social and Behavioral Sciences*. 2014. Vol. 138. P. 584-591. 11. *Giaglis G. M., Minis I., Tatarakis A., Zeimpekis V.* Minimizing logistics risk through real-time vehicle routing and mobile technologies: research to date and future trends // *International Journal of Physical Distribution & Logistics Management*. 2004. Vol. 34. P. 749-764. 12. *Victor Pillac, Michel Gendreau, Christelle Gureret, Andries Medaglia.* A review of dynamic vehicle routing problems // *European Journal of Operational Research, Elsevier*. 2013. Vol. 225 (1). P. 1-11. 13. *Петров Е.Г., Новожилова М.В., Гребенник І.В.* Методи і засоби прийняття рішень у соціально-економічних системах. К.: Техніка. 2004. 256 с. 14. *Чуб О.І., Новожилова М.В.* Оптиміальний розподіл ресурсів при реалізації проєктів реконструкції інженерних мереж в мультипроєктному середовищі // *Вісник НТУ «ХПІ»*. 2013. Вип. 6. С. 67-75. 15. *Чуб І.А., Іванілов А.С., Новожилова М.В.* Решение задачи распределения ресурсов проєкта как оптимизационной задачи размещения геометрических объектов с изменяемыми метрическими характеристиками // *Проблемы машиностроения*. 2010. Т. 3, № 5-6. С. 79-90.

Транслітерований список літератури: 1. *Mirotin L.B., Gudkov V.A., Zyrjanov V.V.* i dr. *Upravlenie gruzovymi potokami v transportno-logisticheskikh sistemah* / Pod red. L.B. Mirotina. M.: Gorjachaja linija-Telekom. 2010. 704 p. 2. *Dantzig G.B., Ramser J.H.* The Truck Dispatching Problem // *Management Science*. 1959. Vol. 6. No. 1. P. 80-91. 3. *Gryshko S.V., Guca O.M., Suhomlinov A.I.* Modeljvannja vyrobnychoj logistyky v umovah perebudovy pidpryjemstva // *Radioelektronika ta informatyka*. 2013. № 2. S. 31-34. 4. *Gvozdyn's'kyj A.M., Obizna M.Ju.* Doslidzhennja zadach transportnogo typu // *Radioelektronika ta informatyka*. 2013. № 4. S. 35-39. 5. *Zadorov V.B., Fedusenko O.V., Fedusenko A.O.* Zastosuvannja metodiv bagatokryterial'noi' optyimizacii' do planuvannja vantaznyh perevezen' // *Upravlinnja rozvytkom skladnyh*

system. 2010. №2. S. 27-31. 6. *Melamed I.I., Sergeev S.I., Sigal I.H.* Zadacha kommivojazhera. Voprosy teorii // *Avtomatika i telemehanika*. 1989. № 9. С. 3-33. 7. *Hajghasem M., Shojaie A.A.* Optimal routing in supply chain aimed at minimizing vehicle cost and supply // *Procedia Economics and Finance*. 2016. Vol. 36. P. 353-362. 8. *Szucs G.* Decision support for route search and optimum finding in transport networks under uncertainty // *Journal of Applied Research and Technology*. 2015. Vol. 13. P. 125-134. 9. *Kuznecov K.A., Gromov V.A.* Podhod k resheniju zadachi optyimizacii struktury distrib'jutorskoj kompanii // *Visnyk AMSU. Serija: «Tehnichni nauky»*. 2014. № 1 (51). S. 94-103. 10. *Lang Z., Yao E., Hu W., Pan Z.* A vehicle routing problem solution considering alternative stop points // *Procedia-Social and Behavioral Sciences*. 2014. Vol. 138. P. 584-591. 11. *Giaglis G. M., Minis I., Tatarakis A., Zeimpekis V.* Minimizing logistics risk through real-time vehicle routing and mobile technologies: research to date and future trends // *International Journal of Physical Distribution & Logistics Management*. 2004. Vol. 34. P. 749-764. 12. *Victor Pillac, Michel Gendreau, Christelle Gureret, Andries Medaglia.* A review of dynamic vehicle routing problems // *European Journal of Operational Research, Elsevier*. 2013. Vol. 225 (1). P. 1-11. 13. *Petrov E.G., Novozhylova M.V., Grebennik I.V.* Metody i zasoby pryjnattja rishen' u social'no-ekonomichnyh sistemah. K.: Tehnika. 2004. 256 s. 14. *Chub O.I., Novozhylova M.V.* Optymal'nyj rozpodil resursiv pry realizacii' proektiv rekonstrukcii' inzhenernyh merezh v mul'typroektному середовищі // *Visnyk NTU «HPI»*. 2013. Vyp. 6. S. 67-75. 15. *Chub I.A., Ivanilov A.S., Novozhilova M.V.* Reshenie zadachi raspredelenija resursov proekta kak optyimizacionnoj zadachi razmeshhenija geometricheskikh ob'ektov s izmenjaemyimi metriceskimi karakteristikami // *Problemy mashinostroenija*. 2010. Т. 3, № 5-6. S. 79-90.

Надійшла до редколегії 04.03.16

Рецензент: д-р фіз.-мат. наук, проф. Сізова Н.Д.

Новожилова Марина Володимирівна, д-р фіз.-мат. наук, професор, завідувач кафедри економічної кібернетики та інформаційних технологій Харківського національного університету будівництва та архітектури. Наукові інтереси: системний аналіз, математичне моделювання складних динамічних систем. Адреса: Україна, 61002, Харків, вул. Сумська, 40, тел. (057) 706-20-49.

Литвиненко Євген Миколайович, доцент кафедри економічної кібернетики та інформаційних технологій Харківського національного університету будівництва та архітектури. Наукові інтереси: інформаційні технології, багатокритеріальна оптимізація. Адреса: Україна, 61002, Харків, вул. Сумська, 40, тел. (057) 700-02-46.

Marina V. Novozhilova, doctor of physical and mathematical sciences, professor, head of department of economic cybernetics and information technologies Kharkiv national university of construction and architecture. Scientific interests: system analysis, mathematical modeling of complex dynamical systems. Address: Ukraine, 61002, Kharkiv, Sumskaia street, 40, Tel.: (057) 706-20-49.

Yevhen M. Lytvynenko, associate professor of the department of economic cybernetics and information technologies Kharkiv national university of construction and architecture. Scientific interests: information technology, multicriteria optimization. Address: Ukraine, 61002, Kharkiv, Sumskaia street, 40, Tel.: (057) 700-02-46.

ПРИНЯТИЕ РЕШЕНИЙ В СИСТЕМЕ КОНТРОЛЯ И АНАЛИЗА ПАРАМЕТРОВ ДВИЖЕНИЯ АВТОТРАНСПОРТА

ПРОХОРОВ В. П., ПРОХОРОВ А. В.

Рассматриваются проблемы автоматизации процесса принятия решений в информационно-навигационных системах контроля и анализа параметров движения автотранспорта на основе интеллектуальных информационных технологий. Показывается возможность применения для этих целей интеллектуальных систем поддержки принятия решений. Приводится структура, принципы построения и решаемые задачи систем поддержки принятия решений. Раскрываются основные режимы функционирования. Приводится эффективный алгоритм логического вывода при различных стратегиях сокращения перебора, реализованный в системе и повышающий эффективность решения логико-аналитических задач и принимаемых решений в информационно-навигационных системах контроля и анализа параметров движения автотранспорта.

Введение

Необходимость оперативного получения навигационной информации привела к формированию обширной комплексной прикладной области транспортно-диспетчерских информационных технологий. Навигационная система мониторинга транспорта или автоматического определения местоположения транспортных средств (Automatic Vehicle Location, AVL) – это программно-аппаратный комплекс, предназначенный для сбора различной информации с парка транспортных средств, в том числе для отслеживания их местоположения, сбора и предоставления этой информации в удобном для анализа виде сотрудникам диспетчерских пунктов, информационно-аналитических центров

(ИАЦ) различного назначения, в региональные, отраслевые АСУ дорожного движения (АСУ ДД). Высокоточное определение навигационных параметров транспортных средств обеспечивается спутниковыми радионавигационными системами или интегрированными навигационными системами. Современные системы мониторинга не ограничиваются отражением маршрутов движения объектов. Они представляют собой комплексный продукт, который способен решить множество вопросов, связанных с логистикой, безопасностью и экономикой перевозок.

Создание информационно-навигационных систем контроля и анализа параметров движения автотранспорта (ИНСКА) [1], ориентированных на повышение безопасности и эффективности транспортных потоков и пользователей транспорта, возможно на основе интеграции средств автоматизации контроля и управления транспортом, информационных и коммуникационных технологий, глобальных навигационных систем, динамических геоданных и единой информационной среды в транспортную инфраструктуру и транспортные средства (рис. 1).

Перед ИНСКА ставятся задачи анализа обстановки на дорогах и маршрутах движения автотранспорта, сопровождения движущихся объектов, их идентификации, контроля за планами передвижения и отклонения от заданных маршрутов, сопоставления параметров движения требованиям правил дорожного движения, оперативного управления ими при различных дорожных происшествиях и др. При этом возрастающая сложность указанных задач, многообразие возникающих ситуаций в процессе контроля большого числа различных автотранспортных средств, анализа больших объемов разнородных данных мониторинга, а также необходимость оперативного принятия соответствующих решений делает актуальным и необходимым использование в ИНСКА интеллектуальных информационных технологий (ИИТ).



Рис. 1. Информационно-навигационная система контроля и анализа параметров движения автотранспорта

1. Анализ последних исследований и публикаций

Интеллектуальные транспортные системы (ИТС) в мире уже давно стали делом привычным [2], а само понятие является очень широким. Часто под ИТС понимается комплекс взаимосвязанных автоматизированных систем, позволяющих оптимизировать работу транспортной системы, решить задачи управления дорожным движением, мониторинга и управления работой различных видов транспорта [3].

Интеллектуальность систем на транспорте дает возможность получения оперативных решений за короткие сроки, в течение которых человек не в состоянии выработать решение, возможность получения новых решений и накопления опыта с занесением его в базы знаний, а также возможность решения комплексных задач, уровень сложности которых исключает возможность их решения человеком [4].

ИТС рассматривается и как класс транспортных систем, обеспечивающих устойчивое, эффективное, экономичное и безопасное функционирование транспорта за счет придания активным элементам транспортной системы свойств адаптивного (интеллектуального) поведения.

Так на железнодорожном транспорте рассматривают создание интеллектуального подвижного состава и инфраструктуры [5], которые формируются на основе самодиагностируемых объектов, обеспечивающих не только передачу оперативной информации о техническом состоянии, остаточном ресурсе, целесообразности изменения режима работы или необходимости вывода из эксплуатации, но и сокращение удельного энергопотребления и эксплуатационных затрат.

Наибольший интерес с точки зрения ИИТ представляет определение ИТС как комплексной системы оптимизации управления транспортными сетями и средствами в масштабе реального времени, обладающей свойствами адаптивности, ситуационного анализа и планирования (предсказания).

Однако зачастую не учитывается, что ИИТ – это, прежде всего, формальный логико-математический инструментальный на основе знаний, который может быть использован для решения задач с позиций общесистемного подхода к анализу и управлению всеми системами и процессами на транспорте с возможностью адаптации к условиям применения и эксплуатации. И, несмотря на значительные достижения в области ИИТ, о широком использовании реальных интеллектуальных систем на транспорте в настоящее время говорить преждевременно.

2. Постановка задачи исследования

Целью данной работы является применение ИИТ [6], а именно, интеллектуальных систем поддержки принятия решений (СППР) в интересах ИНСКА.

Интеллектуальные СППР предназначены обеспечить лицу, принимающему решения (ЛПР), возможность

анализа ситуаций, постановки задач, выработки, контроля и оценивания вариантов решений, обеспечивающих достижение поставленной цели, а также решать задачи, которые не формализованы или трудно формализуются в предметной области, на основе знаний, предоставленных экспертами, и организовать взаимодействие ЛПР с компьютером в привычных понятиях, терминах, образах.

Целью применения СППР является повышение эффективности работы ЛПР путем улучшения качества, достоверности и сокращения сроков выработки решений с помощью накопленного опыта, адаптации к условиям применения и эксплуатации, учета полноты значимых факторов и оптимизации поиска решений.

3. Интеллектуальные системы поддержки принятия решений для ИАЦ

Основное отличие интеллектуальных СППР – наличие баз знаний (БЗ). Под БЗ понимается упорядоченная совокупность правил, фактов, механизмов вывода и программных средств, описывающих предметную область, предназначенную для представления накопленных в ней знаний [6]. Механизмы (процедуры или алгоритмы) логического вывода (МЛВ) – это совокупность правил вывода и способов их применения этих правил.

Была разработана информационная технология автоматизированного проектирования СППР для ИАЦ и АСУ ДД, которая представляет собой облачную платформу разработки интеллектуальных СППР [7]. Инструментальная среда включает средства проектирования баз знаний и выбора режима работы решателя задач и обеспечивает: поддержку всех этапов по разработке СППР, их адаптацию для решения прикладных задач в ИАЦ и АСУ ДД, в том числе для контроля и анализа параметров движения автотранспорта и других задач, решаемых в ИНСКА. Кроме того, данная платформа обеспечивает хранение баз знаний и данных в датацентрах; предоставление доступа к ней как к сервисам через веб-интерфейс. Интеллектуальные СППР позволяют производить в режиме реального времени получение данных из серверов датацентра или внешних источников, рассуждения на основе правил, хранящихся в БЗ, предоставление результатов для визуализации в естественно-языковой форме для пользователей, архивирование и др. Поддерживается работа с онтологиями и вывод на них, что дает возможность подключения дополнительных категоризированных БЗ. Сервисы платформы обеспечивают наполнение, корректировку и пополнение БЗ СППР в процессе их эволюционного развития.

При создании моделей знаний используется простой и удобный язык описания экспертных знаний. Поддержка различных источников данных и веб-сервисов позволяет проще интегрировать СППР в разнородное программное окружение.

В ИАЦ с использованием СППР могут решаться следующие задачи:

- анализ, оценка и распознавание ситуаций, объектов;
- обобщение результатов распознавания ситуаций, объектов от нескольких источников обнаружения для формирования единого информационного поля;
- контроль, оценка и диагностика состояний, параметров;
- оценка важности и приоритета альтернатив, ситуаций, объектов, решений и т.п.;
- выявление и оповещение о чрезвычайных, аварийных ситуациях, угрожающих состояниях и их регистрация;
- прогнозирование развития событий, ситуаций и действий;
- реализация алгоритмов и сценариев действий;
- выбор наиболее рациональных (оптимальных) планов действий и контроль за их выполнением;
- формирование рекомендаций, советов и оценивание вариантов решений;
- обеспечение быстрой настройки, адаптации и модификации системы к новым условиям эксплуатации.

Основными структурными составляющими СППР являются база знаний и механизм логического вывода. База знаний включает в себя базу правил или аксиом и базу вопросов и предназначена для формализованного описания логических задач на простом внутреннем языке описания экспертных знаний. В основу формализации задач предметной области системы положено логическое исчисление предикатов первого порядка. Модуль (механизм) логического вывода – центральный элемент ядра программного комплекса, предназначенный для логического вывода следствий (ответов) из системы правил (аксиом), которые находятся в БЗ, с использованием модифицированного метода резолюций [8].

Гибкая открытая структура СППР расширяет функциональные возможности системы и круг задач, решаемых в процессе ее эксплуатации, а также постоянно повышает точность анализа, прогнозирования, планирования, организации, координации и контроля принимаемых решений за счет накапливаемого в БЗ опыта. Наличие достаточно полных моделей знаний в конкретной предметной области, например, в системе контроля и анализа параметров движения автотранспорта, и постоянный контроль тенденции изменения параметров объекта управления обеспечивает диагностику и прогноз его поведения с высокой степенью достоверности и заданной точности.

Структура СППР представлена на рис. 2.

В общем случае в БЗ для каждого предиката указывают: его тип (первичный – означает фактами, вторичный – определяется формулой, расчетный – означает фактами, полученными в результате логического вывода); внутреннее представление на специ-

альном формализованном языке; семантическое содержание; формулу вычисления предиката (для вторичных); формулу определения фактов (для расчетных); семантическое содержание рекомендации. Семантическое содержание предиката представляет собой текст ответа на вопрос. В семантическом содержании указывают переменные (куда будут подставлены значения, найденные в результате логического вывода). Кроме указанных, есть еще один тип предикатов – предикаты порядка (БОЛЬШЕ, МЕНЬШЕ, РАВНО и др.), истинность которых вычисляется внутренним программным модулем алгоритма логического вывода.

Предикат – “Имя предиката (V_1, V_2, \dots, V_n)” есть n -местное (зависимое от n переменных) отношение между объектами рассматриваемой предметной области, выделенными экспертом для описания задачи и обозначенными именами – переменными: V_1, V_2, \dots, V_n . При подстановке (интерпретации) вместо переменных (аргументов предиката) конкретных значений предикат становится высказыванием об отношении (свойствах) между этими конкретными объектами со значением истинности – «истина» или «ложно».

4. Автоматизация процесса принятия решений в ИНСКА

Механизм логического вывода в СППР представляет собой модифицированный метод резолюций. В исчислении предикатов первого порядка перед выполнением процедуры доказательства необходимо каждое правило преобразовать к виду так называемой Сколемовской стандартной формы. В итоге любая формула может быть представлена множеством дизъюнктов, т.е. дизъюнкцией литер (литера есть предикат или его отрицание). Иногда дизъюнкт называют предложением. Кванторы общности опускают, предполагая, что все переменные, ими связанные, универсально квантифицированы.

Основная идея принципа резолюции заключается в проверке, содержит ли множество дизъюнктов S пустой дизъюнкт. Когда дизъюнкт не содержит никаких литер, то он называется пустым. Так как пустой дизъюнкт не содержит литер, которые могли бы быть истинными при любых интерпретациях, то он всегда ложен. Если S содержит пустой дизъюнкт, то S противоречиво (невыполнимо). Если S его не содержит, то следующие шаги заключаются в выводе новых дизъюнктов до тех пор, пока не будет получен пустой дизъюнкт (что всегда будет иметь место для невыполнимого S). Таким образом, принцип резолюций рассматривается как правило вывода, с помощью которого порождаются новые дизъюнкты из S .

Если в любых двух дизъюнктах C_1 и C_2 имеется контрарная пара литер, то при вычеркивании ее формируется новый дизъюнкт из оставшихся частей дизъюнктов. Этот вновь сформированный дизъюнкт называется резольвентой дизъюнктов C_1 и C_2 . Обоснованность получения таким образом резольвент вытекает

из теоремы о том, что резольвента S , полученная из двух дизъюнктов $S1$ и $S2$, есть логическое следствие этих дизъюнктов.

Образование резольвент продолжается до тех пор, пока не будет получено пустое предложение – доказательство неудовлетворимости множества предложений. В некоторых случаях при образовании резольвент вместо одних переменных подставляют другие переменные или константы. Операция подстановки называется унификацией. При унификации подставляют термы вместо переменных. Если термы предиката $P1$ (предикат контрарной пары из резольвенты на очередном шаге) и термы предиката $\sim P1$ (предикат из выходного множества дизъюнктов) совпадают, то переходят к формированию очередной резольвенты. Если термы не совпадают и терм в $P1$ есть переменная, то вместо нее подставляют терм из предиката $\sim P1$. Если оба терма – константы (неуспех), то выбирают очередного кандидата. Бывают ситуации, при которых унификация невозможна, т.е. перебор предложений-кандидатов не приводит к успеху, возникают тупиковые ситуации. В этом случае необходимо использовать специальные механизмы для их устранения, например, пошаговый возврат на вышестоящие уровни

образования резольвент для проверки альтернативных вариантов. При этом необходимо запоминать факты образования резольвент.

Пусть в базе знаний записано правило

ОБЪЕКТ($V1, V2$)?ТИП($V1, V3$)?РАВНО($V3, 1$)®ДЕЙСТВИЕ($V2$).

В базе данных содержатся следующие данные:

ОБЪЕКТ(5, TP),

ТИП(5, 100).

Вопрос, задаваемый системе:

ДЕЙСТВИЕ(VX).

После преобразования получим следующее множество дизъюнктов:

~ОБЪЕКТ($V1, V2$)Ъ~ТИП($V1, V3$)ЪРАВНО($V3, 100$)ЪДЕЙСТВИЕ($V2$),

ОБЪЕКТ(5, TP),

ТИП(5, 100),

~ДЕЙСТВИЕ(VX).

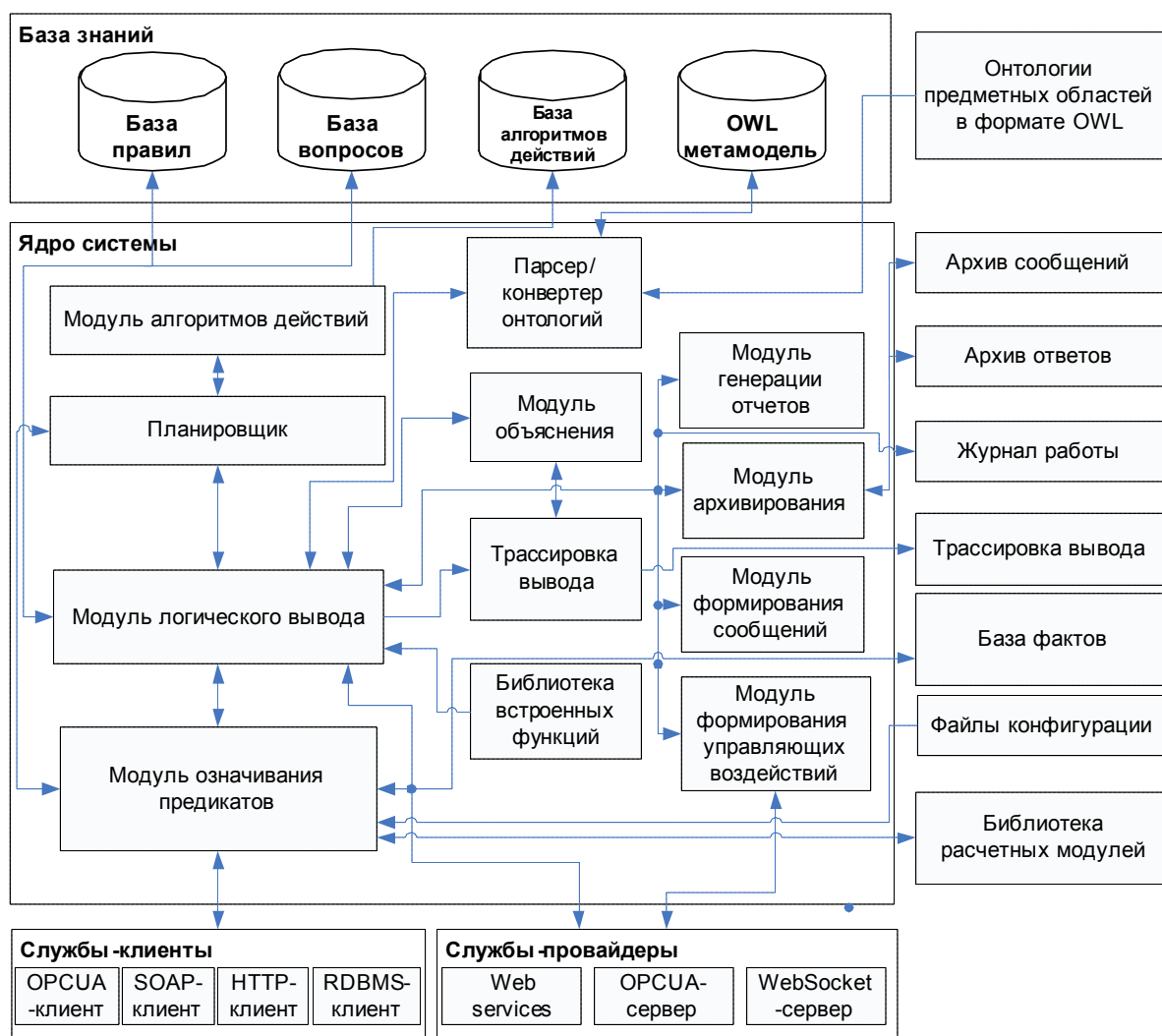


Рис. 2. Структура СППР

Для нахождения ответа к предположению добавляем его отрицание через дизъюнкцию

\sim ДЕЙСТВИЕ(VX) \vee ДЕЙСТВИЕ(VX).

Процесс продолжается до тех пор, пока в корне графа не будет получено предложение, содержащее в качестве термов константы, которые являются решением задачи (рис. 3).

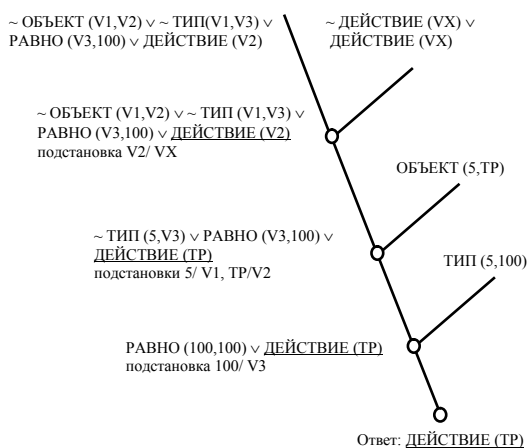


Рис. 3. Пример дерева вывода

Сложность существующих методов доказательства в исчислении предикатов заключается в необходимости перебора большого количества вариантов доказательства в процессе поиска решения. Поэтому большие временные затраты при реализации данных методов сводят к нулю их практическую ценность.

В предлагаемой интеллектуальной системе реализован эффективный алгоритм логического вывода при различных стратегиях сокращения перебора, структурная схема которого показана на рис. 4.

5. Режимы функционирования и возможности СППР

Основные режимы функционирования СППР:

- корректировка и пополнение базы знаний;
- консультирование или вопросно-ответный режим (система выдает ответы – диагнозы, рекомендации, сообщения – на естественном языке в реальном времени принятия решений);
- автоматический (основной режим для автономного варианта применения);
- объяснение результатов (выдается исчерпывающая информация о причинах получения данного ответа – фактов и правил, задействованных в процессе логического вывода – в варианте применения в АРМ).

Функциональные возможности СППР:

- обеспечение быстрой модификации и адаптации к новым условиям функционирования;
- средства интеграции с другими информационными системами и различными СУБД;

- открытость, переносимость и настройка СППР на решаемые задачи в различных предметных областях;

- интерпретация изменяющихся во времени данных, поступающих от внешних источников;

- автоматический вывод и выдача сообщений;

- автоматизированный вывод рекомендаций в виде ответов на вопросы;

- поддержка и сопровождение действий пользователя в различных ситуациях с применением базы алгоритмов или сценариев;

- формирование управляющих воздействий и выдача их во внешние устройства и другие информационные системы;

- архивирование результатов решения логических задач;

- формирование отчетов.

6. Принятие решений в ИАЦ с использованием СППР

Процесс принятия решений в ИАЦ и АСУ ДД с использованием СППР осуществляется на основе анализа обстановки, ситуаций, определяемых и связанных с параметрами движения автотранспорта, правилами дорожного движения, реальными условиями дорожного движения (состоянием дорожного покрытия, наличием дорожных знаков и разметки, погодными условиями, состоянием автотранспорта и водителя и др.). В процессе логического вывода в СППР формируются рекомендации по принятию решений в конкретной дорожной обстановке. На рис. 5 приведена общая схема формирования таких решений в СППР. Данная схема наглядно показывает этапы интерпретации знаний на основе конкретных данных и фактов в складывающейся обстановке дорожного движения, связанной с контролируемым автотранспортом.

Заключение

Использование СППР в ИАЦ и АСУ ДД обеспечивает следующие преимущества:

- облегчает разработку, тестирование, развертывание и сопровождение прикладных задач;
- повышает степень интеллектуализации и адаптации существующих ИАЦ и АСУ ДД к изменению задач и целей функционирования;
- создает простые условия без дополнительного программирования для накопления и повторного использования знаний;
- повышает качество, достоверность и сокращает время выработки и принятия решений в различных режимах применения;
- повышает гибкость и эффективность процессов контроля, принятия решений, управления и функционирования АСУ ДД в различных ситуациях и сложных условиях дорожного движения.

Таким образом, разработанная облачная платформа обеспечивает поддержку всех этапов по созданию интеллектуальных СППР и адаптации их для решения прикладных задач в целях ИАЦ и АСУ ДД. Создаваемые СППР обеспечивают решение задач анализа, диагностики, оценивания и распознавания ситуаций, прогнозирования развития событий, помогают выбрать наилучшие решения в зависимости от возникшей ситуации, сопровождают этапы реализации решений, формируют управляющие воздействия, корректируя тем самым маневры объектов при выполнении задания и оптимизируя его параметры по заданным критериям.

Гибкая открытая архитектура системы обеспечивает быструю настройку, адаптацию и модификацию к новым задачам и условиям эксплуатации. Реализован эффективный алгоритм логического вывода при различных стратегиях сокращения перебора. Примененный подход позволяет не просто создать обособленную экспертную систему обработки данных мониторинга, но и интегрировать компоненты интеллектуального анализа и логического вывода в любую существующую или проектируемую систему контроля и анализа параметров движения автотранспорта.

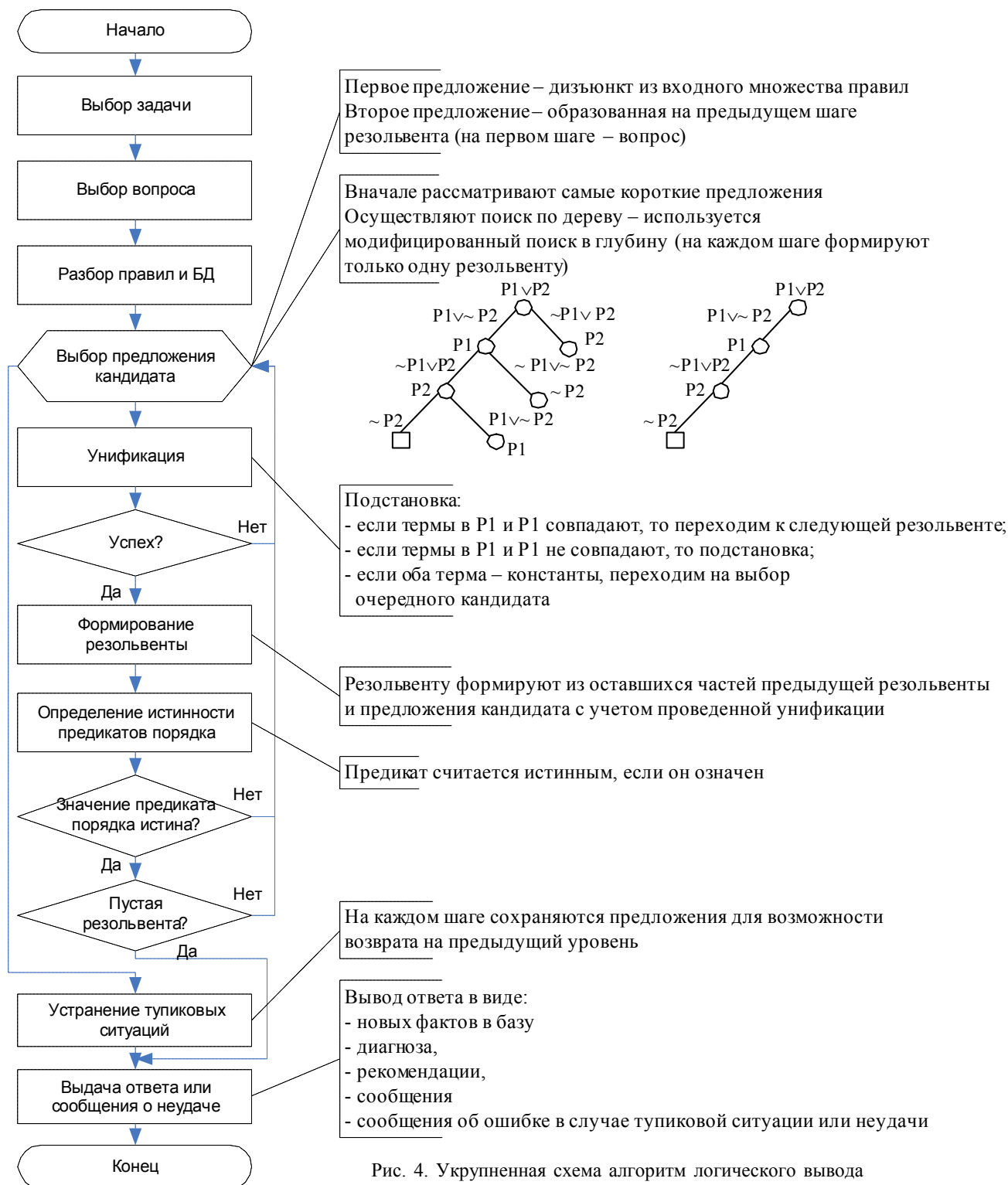


Рис. 4. Укрупненная схема алгоритм логического вывода

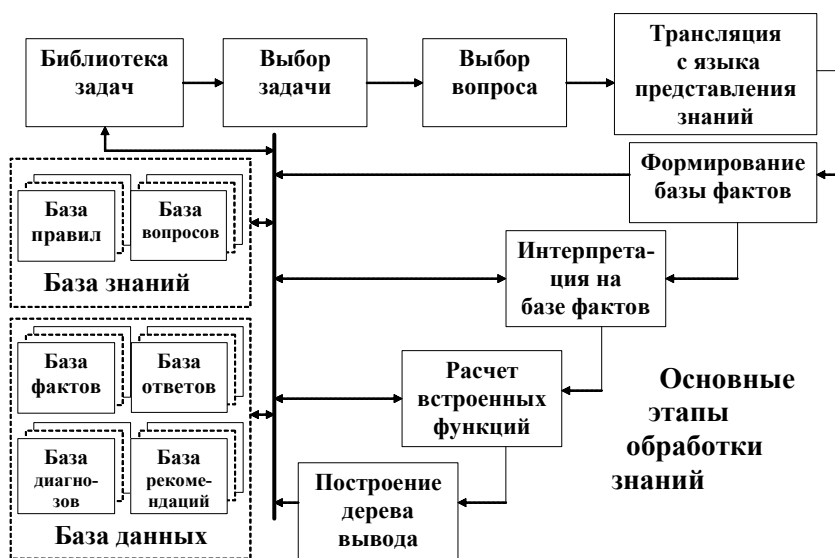


Рис. 5. Общая схема формирования решения в СППР

Сказанное выше определяет *научную новизну и практическую значимость* полученных результатов.

Литература: 1. Дохов А. И. Автоматизированный контроль соблюдения правил дорожного движения. Технические, экономические и правовые аспекты / А. И. Дохов, А. М. Лукьянов, Е. В. Гринченко // Вестник ХНАДУ, 2013. №61-62. С. 110-115. 2. Рейцен Е. А. Перспективы интеллектуальных транспортных систем в городах Украины / Е. А. Рейцен, А. В. Толок, Э. В. Руденков // Материалы XVIII международной научно-практической конференции «Социально-экономические проблемы развития и функционирования транспортных систем городов и зон их влияния». 2012. Т. 5. С. 15-26. 3. Маркелов В. М. Интеллектуальные транспортные системы как инструмент управления / В. М. Маркелов, И. В. Соловьев, В. Я. Цветков // Госсотвестник, 2014. №3 (7). С. 42-47. 4. Боровской А. Е. Внедрение интеллектуальных транспортных систем в рамках национальных программ повышения безопасности дорожного движения / А. Е. Боровской, И. А. Новиков, А. Г. Шевцова // Вестник ХНАДУ, 2013. №61-62. С. 279-283. 5. Интеллектуальные транспортные системы железнодорожного транспорта (основы инновационных технологий) Пособие / В. В. Скалозуб, В. П. Соловьев, И. В. Жуковицкий, К. В. Гончаров. Д. : Изд-во Днепрпетр. нац. ун-та ж.-д. трансп. им. акад. В. Лазаряна, 2013. 207 с. 6. ДСТУ 2481-94. Системи оброблення інформації. Інтелектуальні інформаційні технології. Терміни та визначення. Введ. 01-01-1995. К.: УкрНДІССІ, 1994. 7. Прохоров А. В. Облачная платформа разработки интеллектуальных систем поддержки принятия решений [Текст] / А. В. Прохоров, В. П. Прохоров, А. О. Матюшко // Вісник Національ-

ного технічного університету «ХП». Серія: Системний аналіз, управління та інформаційні технології. 2015. Х.: НТУ «ХП», №58 (1167). С. 72-76. 8. Володин М. И. Определение состава и структуры программного обеспечения СППР для автоматизированных систем управления сложными системами / М. И. Володин, В. П. Прохоров, А. В. Прохоров // Сб. НТ. Системы обработки информации. Х.: ХВУ, 2004. Вып.1. С.187-191.

Поступила в редколлегию 12.02.2016

Рецензент: д-р техн. наук, проф. Федорович О.Е.

Прохоров Валерий Павлович, канд. техн. наук, с.н.с., генеральный директор научного парка «Радиоэлектроника и информатика». Научные интересы: интеллектуальные системы,

системы поддержки принятия решений, механизмы логического вывода, формализованные теории, многозначные логики, категории и морфизмы. Адрес: Украина, 61166, Харьков, пр. Науки, 14, e-mail: nprei.kharkov@gmail.com.

Прохоров Александр Валерьевич, канд. техн. наук, доцент кафедры информационных управляющих систем Национального аэрокосмического университета им. Н. Е. Жуковского «ХАИ». Научные интересы: распределенные интеллектуальные системы, мультиагентные технологии, онтологии, Semantic Web, grid и cloud компьютеринг, промышленная автоматизация, моделирование систем. Адрес: Украина, 61070, Харьков, ул. Чкалова, 17, e-mail: al_val@mail.ru.

Prokhorov Valery, Candidate of Technical Sciences, Senior Research Fellow, CEO of the Science Park «Radio Electronics and Informatics», Kharkiv National University of Radio Electronics. Research interests: intelligent systems, decision support systems, automated reasoning, nonclassical logics, category theory. Address: Ukraine, 61166, Kharkiv, Science Avenue, 14, e-mail: nprei.kharkov@gmail.com.

Prokhorov Alexander, Candidate of Technical Sciences, Associate Professor, Associate Professor of Information Control Systems Department, National Aerospace University «Kharkiv Aviation Institute». Research interests: distributed intelligent systems, multi-agent technology, ontology, Semantic Web, grid and cloud computing, industrial automation, systems modeling. Address: Ukraine, 61070, Kharkiv, Str. Chkalov, 17, e-mail: al_val@mail.ru.



ПОХИБКА ТРИВИМІРНОЇ РЕКОНСТРУКЦІЇ ПОВЕРХНІ ЗА ТРИАДОЮ ЇЇ ЗОБРАЖЕНЬ

ГРАБОВСЬКА Н.Р., РУСИН Б.П., ІВАНЮК В.Г.

Розглядається метод тривимірної реконструкції поверхні за тріадою двовимірних зображень, які отримані під різними напрямками освітлення. Оцінка похибки реконструкції похідних поверхні проводиться для незапланованих відхилень вертикального напрямку освітлення.

1. Вступ

Останнім часом для опису та відновлення форми дефектів, які утворилися на поверхні конструкцій в процесі експлуатації, використовують методи тривимірної реконструкції [1-9]. Одним з методів отримання тривимірної форми предмета на основі аналізу його зображень є запропонований в [10, 11] метод 3D реконструкції за тріадою зображень, отриманих зі змінним положенням джерела світла. Метод оснований на ламбертівській моделі відбиття світла від поверхні конструкції. Для відеозйомки зображень використовуються три напрямки освітлення - вертикальний та два бокові.

Сама реконструкція відбувається шляхом інтегрального накопичення реконструйованих горизонтальної та вертикальної похідної вектора нормалі. Як правило, визначення глибини дефекту відбувається з певною похибкою. Похибка має дві складові: перша, яка виникає лише в розрахунковій частині алгоритму реконструкції, і друга, обумовлена неточністю відеозйомки зображень. Спотворення зображень викликане неточністю встановлення напрямків освітлення поверхні дефекту. За результатами роботи [11] похибка обрахунку поверхні дефекту, яка зосереджена лише в розрахунковій частині алгоритму реконструкції за тріадою зображень для випадку оцінки глибини тріщини, становить $\pm 0,04\%$. Але при практичній реалізації цієї реконструкції необхідно врахувати похибки, які індуковані відхиленням кожного з трьох напрямків освітлення від технологічно запланованого значення. Це складне завдання. Тому, щоб досягти його успішного виконання, слід розбити вирішення цієї проблеми на простіші етапи і аналізувати зазначені похибки по частинах. На першому етапі такого аналізу точності розглянуто похибку, яка індукована відхиленням кожного з двох бокових напрямків освітлення від технологічно запланованого значення у випадку дослідження тріщини [12].

На другому етапі такого аналізу точності доцільно розглянути похибку, яка індукована відхиленням вертикального напрямку освітлення від технологічно запланованого значення.

Тому метою роботи є оцінка похибки визначення похідних поверхні при її реконструкції за тріадою зображень залежно від вертикального напрямку освітлення.

Для досягнення поставленої мети треба вирішити дві задачі оцінки похибок 3D реконструкції похідних поверхні залежно від напрямку освітлення, а також з'ясувати, в якій мірі цей тип похибок можна контролювати в системі тривимірної реконструкції, що дозволить конструктивно зменшити їх до необхідного рівня і покращити точність 3D реконструкції.

2. Ілюмінаційна модель зображення

Для дифузної моделі відбиття (розглядають монохроматичне зображення) інтенсивність відбиття точки поверхні, освітленої одним джерелом світла, визначає рівняння [13, 14]

$$I_{\lambda} = I_a + I_{p\lambda} k_d (N \cdot L), \quad (1)$$

де I_a – інтенсивність розсіяного світла; $I_{p\lambda}$ – інтенсивність джерела освітлення; k_d – дифузний коефіцієнт, який визначає рівень дифузного відбиття досліджуваної точки поверхні; $L = (L_x, L_y, L_z)$ – напрям на джерело світла; $N = (N_x, N_y, N_z)$ – одиничний вектор нормалі до поверхні $F(x, y, z) = 0$ у досліджуваній точці $M = (x, y, z)$.

Вважаємо, що наступну реконструкцію виконують, коли поверхня задана в явній формі $z = f(x, y)$, тому $F(x, y, z) = f(x, y) - z = 0$. Направлені косинуси нормалі (N) до такої поверхні в точці $M = (x, y, z)$ описуються формулами [15]:

$$N = \left\{ \frac{-p}{\sqrt{p^2 + q^2 + 1}}, \frac{-q}{\sqrt{p^2 + q^2 + 1}}, \frac{1}{\sqrt{p^2 + q^2 + 1}} \right\}, \quad (2)$$

де $p = \frac{\partial z}{\partial x}$, $q = \frac{\partial z}{\partial y}$ – нахил поверхні в околі точки.

Оскільки інтенсивність джерела освітлення $I_{p\lambda}$ є стала величина, то рівняння (1) є функцією чотирьох невідомих p , q , k_d та I_a .

3. 3D реконструкція за тріадою зображень

Розглянемо ілюмінаційну модель зображення з контрольованим кутом освітлення та визначимо її параметри p , q , k_d та I_a з тріади зображень. Щоб усунути залежність від параметра I_a , експериментально визначимо його, тимчасово штучно затінюючи зразок з дефектом. Затінення формується за допомогою прямокутника, просторове положення якого над зразком, а отже і положення тіні, визначає система оброб-

ки. Використовуючи базове зображення I_λ зі штучним затіненням, можна визначити інтенсивність фону I_a і звести модель (1) до трьох невідомих.

Забезпечивши у системі джерело світла з такими параметрами, що

$$L_0 = (L_{x0}, L_{y0}, L_{z0}), \quad L_1 = (L_{x1}, L_{y1}, L_{z1}), \\ L_2 = (L_{x2}, L_{y2}, L_{z2}),$$

отримаємо інтенсивність відбиття поверхні у вигляді системи трьох рівнянь:

$$I_{\lambda 0} = I_a + I_{p\lambda} k_d (L_{x0} N_x + L_{y0} N_y + L_{z0} N_z), \\ I_{\lambda 1} = I_a + I_{p\lambda} k_d (L_{x1} N_x + L_{y1} N_y + L_{z1} N_z), \quad (3) \\ I_{\lambda 2} = I_a + I_{p\lambda} k_d (L_{x2} N_x + L_{y2} N_y + L_{z2} N_z),$$

в якій невідомі компоненти вектора нормалі N . Кожна компонента вектора нормалі N залежать від похідних p, q , а отже система (3) залежать від них. Щоб спростити вираз (3), застосуємо у системі джерела світла з такими параметрами, що

$$L_0 = (0, 0, 1), \quad (4)$$

$$L_1 = (L_{x1e}, 0, L_{z1e}), \quad L_2 = (0, L_{y2e}, L_{z2e}). \quad (5)$$

Представимо джерела світла (5) з застосуванням кутів (рис. 1)

$$L_1 = (\cos\theta, 0, \sin\theta), \quad L_2 = (0, \cos\phi, \sin\phi). \quad (6)$$

Крім того, приймаємо, що джерела світла L_{S0}, L_{S1}, L_{S2} є однакові за інтенсивністю.

Отриману систему трьох рівнянь з невідомими параметрами k_d та p, q шляхом алгебраїчних перетворень трансформуємо у таку форму:

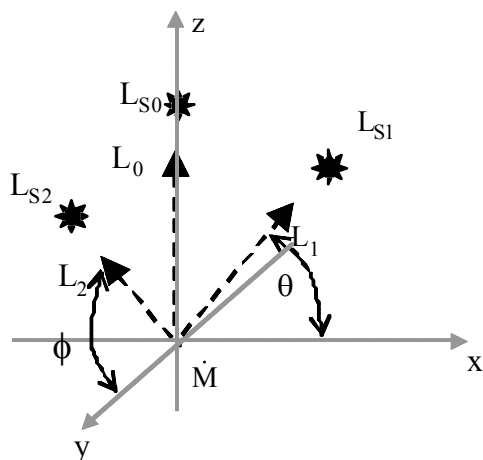


Рис. 1. Геометрична інтерпретація локалізації джерела світла (L_{S0}, L_{S1}, L_{S2}) в тривимірному просторі відносно точки \dot{M} поверхні

$$p = \frac{L_{z1e}}{L_{x1e}} - \frac{\eta_1}{L_{x1e}}, \quad (7)$$

$$\text{де } \eta_1 = \frac{I_{\lambda 1} - I_a}{I_{\lambda 0} - I_a},$$

$$q = \frac{L_{z2e}}{L_{y2e}} - \frac{\eta_2}{L_{y2e}}, \quad (8)$$

$$\text{де } \eta_2 = \frac{I_{\lambda 2} - I_a}{I_{\lambda 0} - I_a}.$$

Оскільки напрямки освітлення L_0 (4), L_1, L_2 (5) є технологічно заплановані, то на основі їх компонент виконується реконструкція (7) і (8).

4. Похибка визначення похідних поверхні

На практиці під час зйомки виникають відхилення від напрямків освітлення (4), (5). Тому на першому етапі аналізу точності реконструкції на основі тріади 2D зображень доцільно розглянути похибку, яка індукована відхиленням напрямку освітлення

$$\tilde{L}_0 = (\tilde{L}_{x0}, \tilde{L}_{y0}, \tilde{L}_{z0}) \quad (9)$$

від технологічно запланованого L_0 для відеозйомки зображення $I_{\lambda 0}$. Зауважимо, що за внаслідок змін напрямку освітлення \tilde{L}_0 зображення $I_{\lambda 0}(L_0)$ спотворюється і стає $I_{\lambda 0}(\tilde{L}_0)$. Тому реконструкція похідної p (7) змінюється лише за рахунок зміни параметра $\eta_1(I_{\lambda 0}(\tilde{L}_0))$, а реконструкція похідної q (8) – лише за рахунок зміни параметра $\eta_2(I_{\lambda 0}(\tilde{L}_0))$.

Розглянемо похибку реконструкції похідної p , яка індукована відхиленням напрямку освітлення \tilde{L}_0 (9).

Для векторів L_0, \tilde{L}_0 скористаємось описом, який слідує при використанні поверхні сфери $x^2 + y^2 + z^2 = 1$, з центром у досліджуваній точці O . Досліджувана точка розміщена у центрі координат.

Для поверхні $z = \sqrt{1 - x^2 - y^2}$ можна визначити похідну $P = \frac{\partial z}{\partial x} = -\text{ctg}\theta$, де θ – кут напрямку світла в площині XOZ , а також можна визначити похідну $Q = \frac{\partial z}{\partial y} = -\text{ctg}\phi$, де ϕ – кут напрямку світла в площині YOZ . На основі цього опис вектору \tilde{L}_0 набуває форми

$$\tilde{L}_0 = \left(\frac{-P_0}{\sqrt{P_0^2 + Q_0^2 + 1}}, \frac{-Q_0}{\sqrt{P_0^2 + Q_0^2 + 1}}, \frac{1}{\sqrt{P_0^2 + Q_0^2 + 1}} \right). \quad (10)$$

Привідсутності відхилення освітлення від технологічно запланованого у точці $P_0 = P_{e0} = 0$, $Q_0 = Q_{e0} = 0$ вираз (7) приймає значення

$$p = p(0, 0) = \frac{L_{zle}}{L_{xle}} - \frac{\eta_l(0, 0)}{L_{xle}}. \quad (11)$$

У випадку, коли маємо відхилення освітлення від технологічно запланованого, зображення $I_{\lambda 0}$ стає функцією параметрів P_0, Q_0 . Внаслідок цього вираз (11) приймає значення

$$p(P_0, Q_0) = \frac{L_{zle}}{L_{xle}} - \frac{\eta_l(P_0, Q_0)}{L_{xle}}. \quad (12)$$

Застосуємо для оцінки точності реконструкції параметра p розклад функції (12) в ряд Маклорена:

$$\begin{aligned} p(P_0, Q_0) &= p(0, 0) + P_0 \left[\frac{\partial p(P_0, 0)}{\partial P_0} \right]_{P_0=0} + \\ &+ Q_0 \left[\frac{\partial p(0, Q_0)}{\partial Q_0} \right]_{Q_0=0} + P_0^2 \frac{1}{2!} \left[\frac{\partial^2 p(P_0, 0)}{\partial P_0^2} \right]_{P_0=0} + \\ &+ Q_0^2 \frac{1}{2!} \left[\frac{\partial^2 p(0, Q_0)}{\partial Q_0^2} \right]_{Q_0=0} + \\ &+ \frac{1}{2!} P_0 Q_0 \left[\frac{\partial^2 p(P_0, Q_0)}{\partial P_0 \partial Q_0} \right]_{P_0=0, Q_0=0} + \dots \end{aligned} \quad (13)$$

Наступний аналіз точності виконується з технологічно запланованим напрямком L_1 , який визначено в точці з $P = P_{e1}$, $Q = Q_{e1} = 0$ таким виразом

$$L_1 = \left(\frac{-P_{e1}}{\sqrt{P_{e1}^2 + 1}}, 0, \frac{1}{\sqrt{P_{e1}^2 + 1}} \right). \quad (14)$$

Для оцінки другого і третього членів розкладу (13), підставивши вирази (2) та (14) у вираз (12), шляхом його диференціювання отримуємо

$$\left[\frac{\partial p(P_0, 0)}{\partial P_0} \right]_{P_0=0} = -\frac{p^2 P_{e1} + p}{P_{e1}}, \quad (15)$$

$$\left[\frac{\partial p(0, Q_0)}{\partial Q_0} \right]_{Q_0=0} = -\frac{q P_{e1} + q}{P_{e1}}. \quad (16)$$

Нехтуючи у виразі (13) малими членами, отримуємо наближення розвинення в ряд Маклорена, з якого, з використанням підставлення виразів (15) і (16) в (13), визначимо загальну оцінку точності реконструкції горизонтальної похідної

$$\Delta p_{PQ} = p(P_0, Q_0) - p(0, 0) = \Delta p_P + \Delta p_Q, \quad (17)$$

де похибки

$$\Delta p_P = P_0 \left[\frac{\partial p(P_0, 0)}{\partial P_0} \right]_{P_0=0}, \quad (18)$$

$$\Delta p_Q = Q_0 \left[\frac{\partial p(0, Q_0)}{\partial Q_0} \right]_{Q_0=0}. \quad (19)$$

Для оцінки точності реконструкції параметра q поступимо аналогічно, розклавши функцію $q(P_0, Q_0)$ в ряд Маклорена:

$$\begin{aligned} q(P_0, Q_0) &= q(0, 0) + P_0 \left[\frac{\partial q(P_0, 0)}{\partial P_0} \right]_{P_0=0} + \\ &+ Q_0 \left[\frac{\partial q(0, Q_0)}{\partial Q_0} \right]_{Q_0=0} + P_0^2 \frac{1}{2!} \left[\frac{\partial^2 q(P_0, 0)}{\partial P_0^2} \right]_{P_0=0} + \\ &+ Q_0^2 \frac{1}{2!} \left[\frac{\partial^2 q(0, Q_0)}{\partial Q_0^2} \right]_{Q_0=0} + \\ &+ \frac{1}{2!} P_0 Q_0 \left[\frac{\partial^2 q(P_0, Q_0)}{\partial P_0 \partial Q_0} \right]_{P_0=0, Q_0=0} + \dots \end{aligned} \quad (20)$$

Наступний аналіз точності виконується з технологічно запланованим напрямком L_2 , який визначено в точці з $P = P_{e2} = 0$, $Q = Q_{e2}$ виразом

$$L_2 = \left(0, \frac{-Q_{e2}}{\sqrt{Q_{e2}^2 + 1}}, \frac{1}{\sqrt{Q_{e2}^2 + 1}} \right). \quad (21)$$

Для оцінки другого і третього членів розкладу (20), підставивши вирази (2) та (21) у вираз (20), шляхом його диференціювання отримуємо

$$\left[\frac{\partial q(P_0, 0)}{\partial P_0} \right]_{P_0=0} = -\frac{q P_{e2} + p}{Q_{e2}}, \quad (22)$$

$$\left[\frac{\partial q(0, Q_0)}{\partial Q_0} \right]_{Q_0=0} = -\frac{q^2 Q_{e2} + q}{Q_{e2}}. \quad (23)$$

Нехтуючи у виразі (20) малими членами, отримуємо наближення розвинення в ряд Маклорена, з якого, з використанням підставлення (22), (23), визначимо загальну оцінку точності реконструкції вертикальної похідної

$$\Delta q_{PQ} = q(P_0, Q_0) - q(0, 0) = \Delta q_P + \Delta q_Q, \quad (24)$$

де похибки

$$\Delta q_P = P_0 \left[\frac{\partial q(P_0, 0)}{\partial P_0} \right]_{P_0=0}, \quad (25)$$

$$\Delta q_Q = Q_0 \left[\frac{\partial q(0, Q_0)}{\partial Q_0} \right]_{Q_0=0}. \quad (26)$$

Продовжимо наступний аналіз точності реконструкції похідної r з параметром $P_{e1} = -1$. Підставивши $P_e = -1$ в (15) і (16), отримаємо

$$\left[\frac{\partial p(P_0, 0)}{\partial P_0} \right]_{P_0=0} = -p^2 + p, \quad (27)$$

$$\left[\frac{\partial p(0, Q_0)}{\partial Q_0} \right]_{Q_0=0} = -q P_{e1} + q. \quad (28)$$

На основі (17) з врахуванням (27) і (28) визначимо відносну похибку обрахунку горизонтальної похідної

$$\delta p = \delta p_P + \delta p_Q, \quad (29)$$

де
$$\delta p_P = \frac{P_0}{g_P}, \quad (30)$$

тут
$$g_P = \frac{1}{1-p}, \quad (31)$$

$$\delta p_Q = \frac{Q_0}{g_Q}, \quad (32)$$

тут
$$g_Q = \frac{P}{(1-p)q}. \quad (33)$$

Далі доцільно розглянути точність реконструкції, проаналізувавши похибку реконструкції вертикальної похідної, викликану змінами напрямку $L_0(P_0, Q_0)$. Продовжимо наступний аналіз точності реконструкції похідної q з параметром $Q_{e2} = -1$. Підставимо $Q_{e2} = -1$ в (25), (26) і визначимо відносну похибку обрахунку вертикальної похідної

$$\delta q = \delta q_Q + \delta q_P, \quad (34)$$

де
$$\delta q_Q = \frac{Q_0}{G_Q}, \quad (35)$$

тут
$$G_Q = \frac{1}{1-q}, \quad (36)$$

$$\delta q_P = \frac{P_0}{G_P}, \quad (37)$$

тут
$$G_P = \frac{q}{(1-q)p}. \quad (38)$$

В такій формі похибки формально подібні. Тому необхідно та достатньо провести аналіз точності для однієї з них. Проведемо цей аналіз для похибки δp_Q (32). За умовами тривимірної реконструкції маємо заплановане номінальне значення параметра Q_0 $Q_{0e} = 0$. Нехай задана абсолютна похибка визначення параметра Q_0 як відхилення від запланованого номінального значення $\pm \Delta Q$, тобто має місце

$$-\Delta Q \leq Q_0 \leq \Delta Q. \quad (39)$$

Таким чином, знак похибки δp_Q (32) спільно формується параметрами Q_0, g_Q .

Нехай задано кількісну оцінку точності реконструкції горизонтальної похідної, як умову реконструкції

$$-\delta p_{Q0} \leq \delta p_Q \leq \delta p_{Q0}.$$

Розглянемо таку умову реконструкції з підставленням δp_Q :

$$-\delta p_{Q0} \leq \frac{Q_0}{g_Q} \leq \delta p_{Q0}. \quad (40)$$

Залежно від знаку похибки δp_Q розглянемо чотири випадки формування умови (40).

В першому випадку нехай $Q_0 = \Delta Q$ і $g_Q > 0$. Внаслідок цього права частина умови (40) трансформується у таку умову:

$$\Delta Q \leq g_Q \delta p_{Q0}. \quad (41)$$

У другому випадку нехай $Q_0 = -\Delta Q$ і $g_Q < 0$. Внаслідок цього права частина умови (40) трансформується у таку умову

$$-\Delta Q \geq \delta p_{Q0} g_Q. \quad (42)$$

У третьому випадку нехай $Q_0 = \Delta Q$ і $g_Q < 0$. Внаслідок цього ліва частина умови (40) трансформується в умову (42).

У четвертому випадку нехай $Q_0 = -\Delta Q$ і $g_Q > 0$. Внаслідок цього ліва частина умови (40) трансформується в умову (41).

Таким чином, коли задані відносна похибка δp_{Q0} та абсолютна похибка $\pm \Delta Q$, для встановлення належності точки (p, q) до робочих діапазонів реконструкції необхідно і достатньо визначити параметр $g(p, q)$ і проаналізувати, чи виконується одна з умов (41), (42).

Використовуючи отримані результати (41), (42), (39) та застосовуючи підставлення $P=Q$, отримуємо аналіз точності реконструкції, яка формується похибкою δp_P .

Зауважимо, що оскільки параметр g_P не залежить від похідної q , то на етапі кількісного аналізу виникає можливість простежити результати (41), (42), використовуючи їх візуалізацію, не обмежуючись одним значенням абсолютної похибки визначення параметра $P_0 - \pm \Delta P$, а у діапазоні похибок $D_{\Delta P} = [\pm \Delta P_S, \pm \Delta P_S]$.

Проведемо кількісну оцінку діапазонів реконструкції похідних p, q на основі умов (41) і (42) та абсолютної похибки параметра $Q_0 (P_0)$ для конкретної системи реконструкції і реального пристрою вертикального променя, зображеного на рис. 2.

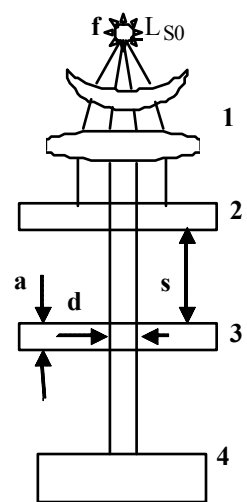


Рис. 2. Пристрій вертикального променя

5. Система тривимірної реконструкції

Щоб з'ясувати практичні аспекти технологічної реалізації точності встановлення напрямку освітлення, розглянемо варіант системи тривимірної реконструкції за тріадою зображень (СТРТЗ) з наступними технічними характеристиками:

1. Відносна похибка реконструкції горизонтальної похідної $\delta p_{P_0} = \pm 0,1$ під впливом параметра P_0 .
2. Відносна похибка реконструкції горизонтальної похідної $\delta p_{Q_0} = \pm 0,1$ під впливом параметра Q_0 .
3. Відносна похибка реконструкції вертикальної похідної $\delta q_{P_0} = \pm 0,1$ під впливом параметра P_0 .
4. Відносна похибка реконструкції вертикальної похідної $\delta q_{Q_0} = \pm 0,1$ під впливом параметра Q_0 .
5. Діапазон реконструкції горизонтальної похідної $D_p = [-0,97; 0,97]$.
6. Діапазон реконструкції вертикальної похідної $D_q = [-0,97; 0,97]$.

Зауважимо, що серед представлених технічних характеристик діапазони D_p, D_q є попередніми характеристиками і їх реальну конфігурацію необхідно кількісно визначити на основі виразів (41), (42).

Розглянемо варіант пристрою вертикального променя (ПВП) в запропонованій СТРТЗ.

5.1. Пристрій вертикального променя

ПВП (див. рис. 2) має такі конструктивні елементи: вузол освітлення, вузол екранування, дослідний зразок 4.

Вузол освітлення складається з джерела світла L_{S0} та конденсора 1. Джерело світла L_{S0} розміщене в фокальній площині конденсора 1, який містить дві лінзи.

Вузол екранування містить два екрани: екран 2 та екран 3. Відстань між екранами s , товщина екрана a .

ПВП має два режими: режим калібровки напрямку освітлення та режим освітлення.

В режимі калібровки встановлюється напрям вертикального освітлення з заданою точністю. В режимі освітлення проводиться освітлення дослідного зразка для потреб тривимірної реконструкції.

Режим калібровки. Вузол освітлення генерує паралельний пучок променів.

Для встановлення вертикального режиму освітлення вузлом освітлення 1 використовуються екрани 2 та 3. Кожний екран має отвір розміром d . Ці отвори технологічно розташовуються на прямій

лінії, яка перпендикулярна до площини дослідного зразка 4.

Вузол освітлення обертом в площині $[XZ]$ (площині $[YZ]$) відносно фокуса f , в якому розміщене джерело світла L_{S0} , встановлюється таким чином, щоб промінь освітлення через отвір екрана 3 падав на дослідний зразок 4. Коли промінь через отвори екранів падає на дослідний зразок 4, то виконується умова перевірки напрямку освітлення. У площині $[XZ]$ ця умова має вигляд

$$\Delta P \geq \frac{d}{s+2a}, \quad (43)$$

а у площині $[XZ]$

$$\Delta Q \geq \frac{d}{s+2a}. \quad (44)$$

ПВП має такі технічні характеристики:

1. Абсолютна похибка $\Delta P_0 = \pm 0,01$ визначення параметра P_0 .
2. Абсолютна похибка $\Delta Q_0 = \pm 0,01$ визначення параметра Q_0 .

Основні габаритні характеристики вузла екранування: $s=30\text{см}$, $a=0,5\text{см}$.

Габаритну характеристику (розмір отвору d) визначимо з умови (43), (44), яка на основі наведених характеристик ПВП приймає таку форму:

$$d \leq 3,1\text{мм}. \quad (45)$$

Приймаємо, що $d=3\text{мм}$.

Зауважимо, що практична реалізація вузла з екранами (див. рис. 2) може бути точно проконтрольована з використанням сучасної вимірювальної техніки, зокрема кутомірів Мікротех.

Режим освітлення. Принцип роботи ПВП в режимі освітлення. Після встановлення кута освітлення отвори екранів збільшуються. Вузол освітлення генерує паралельний пучок променів. Паралельний пучок променів падає на дослідний зразок 4.

6. Кількісна оцінка робочих діапазонів реконструкції горизонтальної похідної

Задаючись технічною характеристикою 1 СТРТЗ, похибкою δp_{P_0} та технічною характеристикою 5 СТРТЗ, діапазоном реконструкції горизонтальної похідної D_p , на основі (41), де застосовано підстав-

лення $P=Q$, розраховано розподіл $\Delta P = P_0 \leq \frac{0,1}{1-p}$.

Результати розрахунків представлені на рис. 3, де $\theta = \arctan p$. Над наведеною кривою (зелений сегмент) реконструкція горизонтальної похідної відбу-

вається з заданою відносною похибкою $\delta p_{p0} = 0,1$.

З рис 3 видно, що при кутах θ , які $87,5^\circ \leq \theta \leq 90^\circ$, реконструкція горизонтальної похідної відбувається з заданою відносною похибкою $\delta p_{p0} = 0,1$ в межах діапазону реконструкції горизонтальної похідної D_p . При куті $87,5^\circ$ абсолютна похибка визначення параметра P_0 $\Delta P = \text{ctg}87,5 = 0,05$, що забезпечує реалізацію технічної характеристики 1 ПВП, абсолютна похибка визначення параметра $P_0 \in \Delta P_0 = \pm 0,01$.

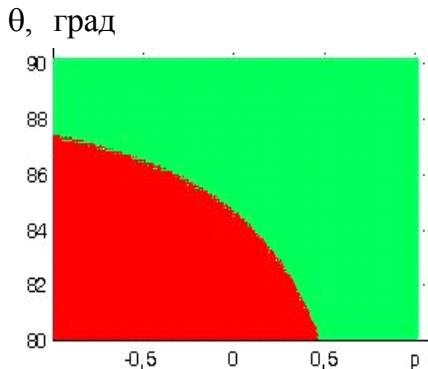


Рис. 3. Залежність кута θ від горизонтальної похідної p

У наступному етапі кількісного дослідження можливості обмежень діапазонів D_p, D_q , задаючись технічною характеристикою 2 СТРТЗ, похибкою δp_{q0} , технічними характеристиками 5, 6 СТРТЗ, діапазонами реконструкції горизонтальної похідної D_p, D_q розрахуємо розподіл

$$Q_0 = 0,1g_Q \quad (46)$$

і спираючись на технічну характеристику 2 ПВП, абсолютну похибку ΔQ_0 , проведемо аналіз виконання умови (41)

$$0,01 \leq Q_0 \quad (47)$$

та умови (42)

$$Q_0 \leq -0,01. \quad (48)$$

Результати розрахунків розподілу Q_0 з кількісним аналізом виконання умов (47) та (48) представлено на рис. 4.

Тут точки (p,q) , в яких можна реконструювати похідну p , коли виконується умова (47), представлено площиною $Q_0 = 0,01$, а точки (p,q) , в яких виконується умова (48), представлено площиною $Q_0 = -0,01$.

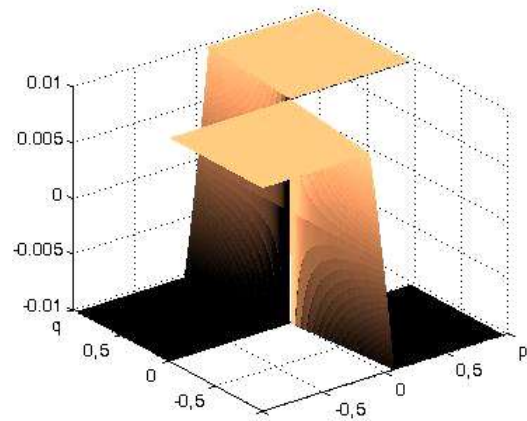


Рис. 4. Частково обмежений розподіл відхилення $Q_0(p,q)$

З рис. 4 видно, що між площинами, де реконструкція виконується з заданою точністю, є зони значень точок (p,q) , де точність реконструкції не відповідає заданій.

Щоб краще селектувати зони реконструкції з заданою точністю і заборонені зони реконструкції, розподіл відхилення Q_0 представлено у бінарній формі на рис. 5.

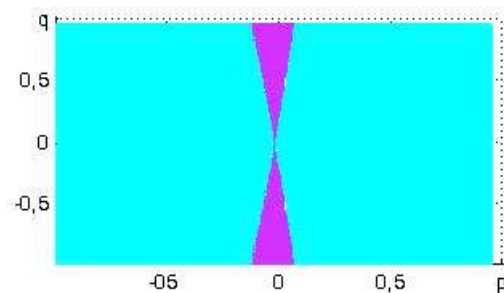


Рис. 5. Бінарний розподіл відхилення $Q_0(p,q)$

На рис. 5 точки, в яких можна реконструювати похідні p та q , представлено сірим кольором, а точки, в яких не можна реконструювати похідні p та q – чорним. З рис. 5, видно, що в околі $p=0$ діапазон реконструкції горизонтальної похідної D_p стає подвійним, а початковий діапазон реконструкції вертикальної похідної D_q звужується. Для кількісної оцінки пошкоджень діапазонів D_p, D_q розподіл похідних p та q в забороненій зоні представлено на рис. 6.

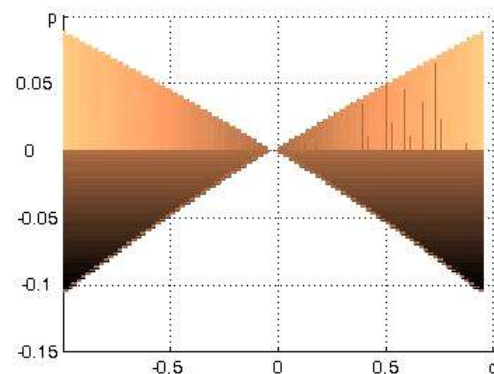


Рис. 6. Сегмент бінарного розподілу $Q_0(p,q)$, в якому є заборонена зона реконструкції

З рис. 6 видно, що діапазон реконструкції горизонтальної похідної D_p стає

$$D_{p-} = [-0,97; p_-], \text{ де } -0,1 \leq p_-(q) < 0,$$

$$D_{p+} = [p_+; 0,97], \text{ де } 0 \leq p_+(q) \leq 0,08.$$

Розширення діапазонів реконструкції відбувається при мінімальному відхиленні похідної q від 0, а екстремальне звуження діапазонів відбувається, коли $q = \pm 0,97$.

Також видно, що коли $-0,1 \leq p \leq 0,08$, початковий діапазон реконструкції вертикальної похідної D_q стає таким: $D_{q-} = [-q_-; q_-]$, де $q_-(p) < 0,97$. Екстремальне звуження діапазону D_{q-} відбувається, коли $p \rightarrow 0$.

Отже, встановлено, що при запланованих технічних характеристиках 1-2 завдяки появі забороненої зони реконструкції заплановані діапазони реконструкції D_p, D_q зменшено.

З виразів (46)-(48) видно, що при фіксованій технічній характеристиці 2 СТРТЗ для розширення діапазонів реконструкції доцільно додатково зменшити абсолютну похибку ΔQ_0 . Один з варіантів практичної реалізації такої операції полягає у збільшенні відстані s . В цьому випадку габарити приладу зростають, а також необхідні додаткові технологічні зусилля для корекції точності розташування напрямків освітлення.

Таким чином, розраховані похибки визначення горизонтальних похідних досліджуваної поверхні обумовлені відхиленням напрямку освітлення L_0 .

Зауважимо, що подібність співвідношень похибок реконструкції (30) та (35), а також співвідношень похибок (32) та (37) дозволяє використати представлені результати з використанням підставлень $q = p$, $p = q$, $P_0 = Q_0$ для кількісної оцінки робочих діапазонів реконструкції вертикальної похідної.

Висновки

Розглянуто задачу тривимірної реконструкції поверхні за тріадою двовірних зображень. Запропоновано алгоритм реконструкції поверхні та розраховано точність його роботи для випадку неконтрольованих змін вертикального напрямку освітлення відеозйомки.

Аналітично знайдені похибки визначення похідних досліджуваної поверхні обумовлені відхиленням вертикального напрямку освітлення від технологічно запланованого значення в площині XOZ та YOZ. На їх основі розраховані похибки визначення горизонтальної та вертикальної похідних поверхні досліджуваного об'єкта, два діапазони 3D реконструкції значень горизонтальної та вертикальної похідних, а також запропоновані рекомендації по встановленню експлуатаційних меж кутових відхилень вертикального напрямку освітлення від технологічно запланованого значення в

перспективі практичної реалізації пристрою вертикального напрямку освітлення.

Література: 1. Y. Liu, X. Su, and Q. Zhang, "A novel encoded-phase technique for phase measuring profilometry," Opt. Express 19(15), 2011. P.14137–14144. 2. Zhang Q., Su X., Xiang L., and Sun X. 3-D shape measurement based on complementary gray-code light, Opt. Lasers Eng. 2012. Vol. 50. Vol. 50 P.574-579. 3. Zhang S., Van Der Weide D., and Oliver J. Superfast phase-shifting method for 3-D shape measurement, Opt. Express. No. 9, 2010. Vol. 18. P. 9684–9689. 4. Woodham R.J. Photometric method for determining surface orientation from multiple images Optical Engineering. 1980. – 19(1). P.139–144. 5. Maria E. Angelopoulou, Maria Petrou Evaluating the effect of diffuse light on photometric stereo reconstruction Machine Vision and Applications. 2014. Vol. 25, Issue 1. P.199-210. 6. Yoshizawa, T. Handbook of Optical Metrology -principle and applications- CRC Press, New York. 2009. 744 p. 7. Zhang S. High-resolution 3-D profilometry with binary phase-shifting methods, Appl. Opt.50(12), 2011 P.1753–1757. 8. Sudipta N. Sinha, D. Scharstein, R. Szeliski. Efficient high-resolution stereo matching using local plane sweeps / IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2014), 2014. P. 1219-1222. 9. Richard Szeliski. Computer Vision: Algorithms and Applications. Springer, New York, 2010. 655P. 10. Неруїнівний контроль стану поверхонь, уражених корозійними пітингами / Б.П. Русин, Н.П. Ануфрієва, Н.Р. Грабовська, В.Г. Іванюк // Фіз.-хім. механіка матеріалів. 2013. №4. С. 90–96. 11. Оцінка глибини тріщини за тріадою зображень / Б.П. Русин, В.Г. Іванюк, О.В. Капшій, Н.П. Ануфрієва // Радіоелектроніка та інформатика. 2010. №2. С. 70–78. 12. Похибка тривимірної реконструкції поверхні тріщини за тріадою зображень / Грабовська Н.Р., Русин Б.П., Іванюк В.Г., Капшій О.В. // Радіоелектроніка та інформатика. 2015. №2. С. 58-63. 13. Lambert J. H. Photometria, sive de Mensura et gradibus luminis, colorum et umbrae / sumptibus viduae E. Klett, 1760. 14. Shintaro Watanabe, Koji Miyajima Detecting Building Changes Using Epipolar Constraint From Aerial Images Taken At Different Positions // ICIP 2001. 2001. P.201-204. 15. Смирнов В.И. Курс высшей математики. Т. 1. М.: Наука. 1967. 480с. 16. tool@microtech-ua.com

Транслітерований список літератури. 1. Y. Liu, X. Su, and Q. Zhang, "A novel encoded-phase technique for phase measuring profilometry," Opt. Express 19(15), 2011. –P.14137–14144. 2. Q. Zhang, X. Su, L. Xiang, and X. Sun, 3-D shape measurement based on complementary gray-code light, Opt. Lasers Eng., vol. 50, 2012, p.574-579. 3. S. Zhang, D. Van Der Weide, and J. Oliver, Superfast phase-shifting method for 3-D shape measurement, Opt. Express vol. 18, No. 9, 2010. P. 9684–9689. 4. Woodham R.J. Photometric method for determining surface orientation from multiple images Optical Engineering. 1980. – 19(1). – P.139–144. 5. Maria E. Angelopoulou, Maria Petrou Evaluating the effect of diffuse light on photometric stereo reconstruction Machine Vision and Applications. 2014, Vol. 25, Issue 1. – P.199-210. 6. Yoshizawa, T. Handbook of Optical Metrology -principle and applications- CRC Press, New York 2009. - 744 p. 7. S. Zhang, High-resolution 3-D profilometry with binary phase-shifting methods, Appl. Opt.50(12), 2011 P.1753–1757. 8. Sudipta N. Sinha, D. Scharstein, R. Szeliski. Efficient high-resolution stereo matching using local plane sweeps. / IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2014), 2014. P. 1219-1222. 9. Richard Szeliski. Computer Vision: Algorithms and Applications. Springer, New York, 2010. 655P. 10. Nerujnivnyj kontrol'

stanu poverhon', urazhenyh korozijnymy pityngamy /B.P. Rusyn, N.P. Anufrijeva, N.R. Grabovs'ka, V.G. Ivanjuk // Fiz.-him. mehanika materialiv. – 2013 –№4 –S. 90–96. 11. Ocinka glybyny trishhyny za triadoju zobrazhen'. / B.P. Rusyn, V.G. Ivanjuk, O.V. Kapshij, N.P. Anufrijeva // Radioelektronika i informatyka. – 2010. – №2–S. 70–78. 12. An error of three-dimensional reconstruction of surface of crack is after the triad of images / Hrabovska N.R., Rusyn B.P., Ivanyuk V.G., Kapshiy O.V. // Radioelektronika i informatyka. 2015. № 2. P. 58-63. 13. J. H. Lambert. Photometria, sive de Mensura et gradibus luminis, colorum et umbrae / sumptibus viduae E. Klett, 1760. 14. Shintaro Watanabe, Koji Miyajima Detecting Building Changes Using Epipolar Constraint From Aerial Images Taken At Different Positions // ICIP 2001. 2001. P.201-204. 15. Smyrnov V.Y. Kurs vjysšej matematyky. T. 1. M.: Nauka. 1967. 480s. 17. tool@microtech-ua.com.

Надійшла до редколегії 22.01.2016

Рецензент: д-р фіз.-мат. наук, проф. Яворський І.М.

Русин Богдан Павлович, д-р техн. наук, проф., зав. відділом “Методи і системи дистанційного зондування “Фізико-механічного інституту ім. Г.В.Карпенка НАНУ. Наукові інтереси: обробка та розпізнавання зображень. Адреса: Україна, 79601, Львів, вул. Наукова, 5а, e-mail: dep32@ipm.lviv.ua.

Грабовська Наталія Романівна, аспірантка Фізико-механічного інституту ім. Г.В.Карпенка НАНУ. Наукові інтереси: обробка та розпізнавання зображень. Адреса: Україна, 79601, Львів, вул. Наукова, 5а, e-mail: dep32@ipm.lviv.ua.

Іванюк Віталій Григорович, інженер відділу “Методи і системи дистанційного зондування “ Фізико-механічного інституту ім. Г.В.Карпенка НАНУ. Наукові інтереси: обробка та розпізнавання зображень. Адреса: Україна, 79601, Львів, вул. Наукова, 5а, тел:2296-530. e-mail: vivan@imp.lviv.ua.

Rusyn B.P., doctor of engineering sciences, professor, manager of department of “Methods and systems of the remote sensing “ of Karpenko Physico- Mechanical Institute of NAS of Ukraine, Address: 79601, Ukraine, Lviv, street Scientific, 5a, telephone : 2296-530, e - mail: dep32@ipm.lviv.ua

Hrabovska N.R., graduate student of Karpenko Physico-Mechanical Institute of NAS of Ukraine, Scientific interests: treatment and artificial perception. Address: 79601, Ukraine, Lviv, street Scientific, 5a, telephone: 2296-530, e - mail: dep32@ipm.lviv.ua

Ivanyuk V.G., engineer of department of “Methods and systems of the remote sensing “ of Karpenko Physico- Mechanical Institute of NAS of Ukraine, . Scientific interests: treatment and artificial perception. Address: 79601, Ukraine, Lviv, street Scientific, 5a, telephone: 2296-530, e - mail: vivan@imp.lviv.ua

КОМПЬЮТЕРНАЯ ИНЖЕНЕРИЯ И ТЕХНИЧЕСКАЯ ДИАГНОСТИКА



УДК004.45

КУБИТНАЯ ФОРМА ОПИСАНИЯ ВЫЧИСЛИТЕЛЬНЫХ СТРУКТУР

TAMER BANI AMER, ЧУМАЧЕНКО С.В.,
ЕМЕЛЬЯНОВ И.В.

Разрабатываются кубитные модели описания структур и функциональностей для повышения быстродействия анализа цифровых устройств за счет увеличения размерности структур данных и памяти. Вводятся основные понятия, термины и определения, необходимые для реализации квантовых вычислений в практику моделирования компьютерных структур. Описываются примеры, подтверждающие эффективность использования кубитных структур данных для параллельного выполнения операций над данными.

Ключевые слова: квантовые векторы, графы и функциональности, кубитные структуры, параллельные вычисления, моделирование цифровых систем.

Введение

Рыночная привлекательность эмуляции квантовых методов вычислений при создании компьютерных структур (КС) в киберпространстве основана на использовании кубитных моделей данных, ориентированных на параллельное решение задач проектирования, тестирования, дискретной оптимизации при существенном повышении затрат памяти. Не вдаваясь в детали физических основ квантовой механики, касающиеся недетерминированного взаимодействия атомных частиц [1, 3], далее используется понятие кубита как двоичного или многозначного вектора для совместного и одновременного задания булеана состояний в дискретной области киберпространства на основе линейной суперпозиции унитарных кодов, ориентированных на параллельное использование методов анализа и синтеза компонентов киберпространства. В быстро развивающейся теории квантовых вычислений векторы состояний образуют квантовый регистр из n кубитов, формирующих унитарное или гильбертово [4, 5] пространство H , размерность которого имеет степенную зависимость от числа кубитов $\text{Dim } H = 2^n$.

Мотивация нового подхода для проектирования КС обусловлена появлением облачных сервисов в рамках новой киберкультуры Internet of Things, которые представляют собой специализированные и рассредоточенные в пространстве виртуальные компьютерные системы [6-9], реализуемые в аппаратуре или в

программном продукте. Любой компонент функциональности, равно как и структура системы, представляется векторной формой, упорядоченной по адресам, таблицей истинности, реализуемой с помощью памяти. Логические функции в традиционном исполнении не рассматриваются. От этого частично уменьшается быстродействие, но, учитывая, что 94% SoC-кристалла составляет память [8], оставшиеся 6% следует реализовывать на памяти, что не будет критичным для большинства облачных сервисов. Практически для создания эффективных компьютерных структур следует использовать теорию, основанную на вычислительных компонентах высокого уровня абстракции: адресуемая память и транзакция.

Особенность организации данных в классическом компьютере состоит в адресации бита, байта или другого компонента. Адресуемость создает проблему обработки ассоциации неадресуемых элементов множества, которые не имеют порядка по определению. Решением может быть процессор, где образ универсума из n унитарно кодированных примитивов использует суперпозицию для формирования булеана $|B(A)|=2^n$ всех возможных состояний [8].

Существует аналогия векторного представления булеана с кубитом квантового компьютера. Квантовому кубиту хранения информации в квантовом компьютере [1], разрешающей суперпозицию $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$, можно поставить во взаимно-однозначное соответствие булеан состояний, формирующий алфавит Кантора $A^k = \{0,1, X, \emptyset\}$, $X = \{0,1\}$. Здесь примитивы алфавита унитарно кодируются векторами: $0 = (10)$ и $1 = (01)$. Коды других символов являются производными по операции суперпозиции $(10) \vee (01) = (11)$ и пересечения $(10) \wedge (01) = (00)$, что формирует булеан: $\{(10), (01), (11), (00)\}$ [6].

Корректность использования прилагательного «кубитная» для моделей цифровых устройств основана на сравнении линейной и булевой алгебры Кантора с алфавитом $A^k = \{0,1, X, \emptyset\}$. Здесь первые два символа – примитивы. Третий определяется суперпозицией: $X = 0 \cup 1$. В гильбертовом пространстве линейная алгебра оперирует примитивами $\alpha|0\rangle$ и $\beta|1\rangle$ кубита, третий символ также есть суперпозиция двух составляющих: $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$. В линейной алгебре нет символа, соответствующего пустому множеству, он есть производная от функции, обратной по отношению к суперпозиции. В теории множеств такой операцией является пересечение, которое дает пустое множество на примитивах: $\emptyset = 0 \cap 1$. В гильбертовом пространстве такой операцией является скалярное произведение или функция Дирака $\langle a|b\rangle$ [4], которая имеет геометрическую интерпретацию:

$$\langle a|b\rangle \approx |a| \times |b| \times \cos \angle(a, b).$$

Если проекции a и b вектора квантового состояния ортогональны, получается:

$$\langle \alpha | \beta \rangle = |a| |b| \cos \angle(a, b) = |a| |b| \cos 90^\circ = 0.$$

Скалярное произведение ортогональных векторов равно нулю, что является аналогом символа пустого множества в алгебре Кантора. Таблица соответствия алгебры множеств и линейной алгебры, представленная ниже, подтверждает свойства изоморфизма между символами булеана и состояниями кубит-вектора:

Boolean $A^k =$	0	1	$X = 0 \cup 1$	$\emptyset = 0 \cap 1$
Qubit $ \psi\rangle =$	$ 0\rangle$	$ 1\rangle$	$\alpha 0\rangle + \beta 1\rangle$	$\alpha 0\rangle \beta 1\rangle$

Следовательно, структуру данных «булеан» можно рассматривать как детерминированный образ квантового кубита в алгебре логики, элементы которой унитарно кодируются двоичными векторами и обладают свойствами суперпозиции, параллелизма и перепутывания. Это дает возможность использовать предлагаемые модели для повышения быстродействия анализа цифровых устройств на классических вычислителях, а также без модификации – в квантовых компьютерах, которые появятся через несколько лет на рынке электроники.

Квантовое описание цифровых функциональных элементов

Кубит (n -кубит) есть векторная форма унитарного кодирования универсума из n примитивов для задания булеана состояний 2^{2^n} с помощью 2^n двоичных переменных. Если $n=2$, то 2-кубит задает 16 состояний с помощью четырех переменных, при $n=1$, кубит задает четыре состояния на универсуме из двух примитивов (10) и (01) с помощью двух двоичных переменных (00,01,10,11) [7]. При этом допускается суперпозиция в векторе 2^n состояний, обозначенных примитивами. Кубит дает возможность использовать параллельные векторные логические операции вместо поэлементных теоретико-множественных для существенного ускорения процессов анализа дискретных систем. Далее кубит отождествляется с n -кубитом или двоичным вектором, если это не мешает пониманию излагаемого материала. Синонимом кубита при задании двоичного вектора логической функции является Q -покрытие (Q -вектор) [8] как унифицированная векторная форма суперпозиционного задания выходных состояний, соответствующих адресным кодам входных переменных функционального элемента. Формат структурного кубитного компонента цифровой схемы $Q^* = (X, Q, Y)$ включает интерфейс (входные и выходную переменные), а также кубит-вектор Q , задающий функцию $Y = Q(X)$, размерность которого определяется степенной функцией от числа входных линий $k = 2^n$. Новизна кубитной формы заключается в замене неупорядоченных по строкам таблиц истинности функциональных элементов векторами упорядоченных состояний выходов.

Если функциональный примитив имеет двоичное покрытие, то ему можно поставить в соответствие кубит или Q -покрытие: $Q=(1110)$:

$P =$	X_1	X_2	Y	$\rightarrow Q = (1110).$
	0	0	1	
	0	1	1	
	1	0	1	
	1	1	0	

Преимущества n -кубита заключается в способности параллельно выполнять логические операции над векторным форматом теоретико-множественных данных. Например, хог-операция над $A=\{a,b,c,d,e,f\}$ и $B=\{a,c,f,g,h,k\}$ выполняется параллельно за один такт, если каждый элемент будет представлен унитарным кодом, а подмножества – векторами, которые являются кубит-операндами:

9-qubit	a	b	c	d	e	f	g	h	k
$A =$	1	1	1	1	1	1	0	0	0
$B =$	1	0	1	0	0	1	1	1	1
$A \oplus B =$	0	1	0	1	1	0	1	1	1

Квантовое описание графовых структур

Квантовый Q -вектор или кубит, естественно, можно и следует использовать для представления графовых структур. Функция и структура – две взаимно обратимые формы описания некоторой сущности. Рассмотренный выше кубит-вектор или квант идеально укладывается в технологию программируемых логических устройств, которые используют адресуемые элементы памяти. Учитывая определенный изоморфизм функций и структур, несложно «запихнуть» в память второй компонент описания дискретной сущности. Для этого необходимо представить некоторый ориентированный граф в виде вершин V , кодированных двоичными векторами и дугами A , создающими направленные переходы между

$$G = (V, A), V = (V_1, V_2, \dots, V_i, \dots, V_n), A = (A_1, A_2, \dots, A_j, \dots, A_n^2).$$

В этом случае ориентированный граф будет представлен таблицей адресов (кодов) всех вершин истоков и стоков, представляющих декартово произведение.

Каждой паре вершин $(V_i, V_j), i, j = \overline{1, n}$ ставится в соответствие 1, если существует дуга между ними $(V_i \rightarrow V_j)$, и 0, если такого соединения нет. Таким образом, квадратичная таблица или матрица смежностей для задания графа превращается в вектор, размерностью n^2 , который технологически тривиально размещается в адресуемом элементе памяти. Такой вектор описания структуры ничем не отличается от задания любой дискретной функциональности и к нему применимы все формальные методы анализа, моделирования, тестирования, верификации и синтеза

дискретных компонентов и систем. Следовательно, создается единый универсальный формат данных в виде кубита или квантового вектора для представления структуры вычислительного устройства и функциональных описаний всех его компонентов. Далее приводятся примеры кубитного задания тривиальных графовых структур, представленных на рис. 1.

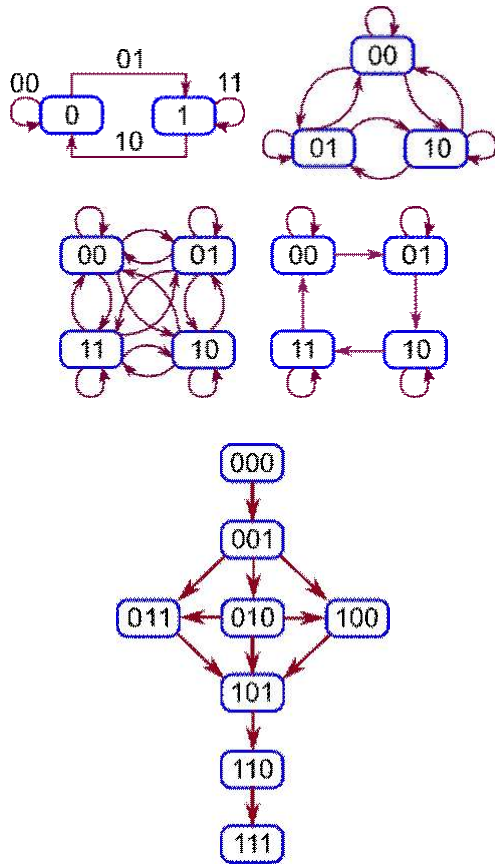


Рис. 1. Примеры примитивных графов для их кубитного описания (G1-G5)

Граф G1 представлен двумя вершинами, которые закодированы 0 и 1 соответственно. Между вершинами имеются четыре перехода, которые обозначены векторами: 00, 01, 10, 11. Данные пары двоичных символов составляют также адреса ячеек памяти, в которые записываются символы 1, поскольку каждый переход имеет место в графе:

$$\begin{array}{|c|c|c|} \hline 0 & 0 & 1 \\ \hline 0 & 1 & 1 \\ \hline 1 & 0 & 1 \\ \hline 1 & 0 & 1 \\ \hline \end{array} \rightarrow \begin{array}{|c|} \hline 1 \\ \hline 1 \\ \hline 1 \\ \hline 1 \\ \hline \end{array} \rightarrow \mathbf{111111} = Q(G_1)$$

Отсутствие некоторого перехода в графе идентифицируется символом 0 в соответствующей ячейке памяти, адрес которой составлен кодами вершин (истока и стока). Граф G2 представлен тремя вершинами, что означает наличие двух двоичных разрядов для кодирования его вершин. Таблица истинности данного графа (повернута для горизонтального инкремента адресов в целях экономии места) также имеет нулевые координаты, поскольку все перехо-

ды, связанные с теоретически возможной вершиной 11, отсутствуют:

$$\begin{array}{|c|c|c|c|c|c|c|c|c|c|c|c|c|c|c|c|} \hline 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ \hline 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ \hline 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ \hline 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ \hline 1 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline \end{array} \rightarrow \mathbf{1110111011100000} = Q(G_2)$$

Если использовать все четыре вершины и построить на них полный граф переходов (G3), то все ячейки кубита, содержащего 16 координат, будут равными 1:

$$\begin{array}{|c|c|c|c|c|c|c|c|c|c|c|c|c|c|c|c|} \hline 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ \hline 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ \hline 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ \hline 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ \hline 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ \hline \end{array} \rightarrow \mathbf{1111111111111111} = Q(G_3)$$

Граф G4 также представлен четырьмя вершинами, но он уже не имеет всех возможных переходов, что означает присутствие нулевых координат в кубитном векторе:

$$\begin{array}{|c|c|c|c|c|c|c|c|c|c|c|c|c|c|c|c|} \hline 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ \hline 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ \hline 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ \hline 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ \hline 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 1 & 1 \\ \hline \end{array} \rightarrow \mathbf{110001100001110011} = Q(G_4)$$

Таким образом, кубитная форма описания графа будет более эффективной, если число вершин в структуре будет стремиться к степени двойки. Кубит следует рассматривать как компактную и неявную форму представления графа, которая однозначно и технологически просто может быть превращена в графический рисунок на основе анализа содержимого адресных ячеек памяти (двоичного вектора).

Вершины графа G5 кодируются уже тремя двоичными разрядами, поскольку число вершин $n=8$. Адрес каждой ячейки кубитного вектора содержит 6 разрядов – удвоенное число разрядов для кодирования вершины, а длина кубит-вектора представлена выражением $q = 2^{2 \times \log_2 n}$. В данном случае число разрядов кубита равно $q=64$, номера координат, в которых находятся единицы: 1, 10, 11, 12, 19, 20, 21, 29, 36, 46, 55, а кубит-вектор имеет следующий вид:

$$Q(G_5) = (0100000001110000001110000001000000100000000010000000100000000100000000)$$

Таким образом, длина кубит-вектора зависит только от количества вершин в графе. Если число дуг мало по сравнению с полным графом, который содержит n^2 направленных соединений, то эффективность E использования кубит вектора – отношение количества единичных координат кубита $\text{card}(1)$ к его общей длине: $E = \text{card}(1) / 2^{2 \times \log_2 n}$ будет невысокой. Например, для графа G5: $E = 11 / 2^{2 \times \log_2 8} = 0,172$.

Анализ кубитного описания графовых структур
Физическая сущность кубита – память (двоичный вектор), в которой каждая адресуемая ячейка хранит значение нуля или единицы. Логическая его сущность – кортеж, состоящий из трех компонентов: (A,B,C),

где A – первая часть адреса ячейки памяти, представленная двоичным кодом вершины истока, B – вторая часть адреса, соответствующая коду вершины стока, C – идентификатор истинности $C(AB)=1$ или ложности $C(AB)=0$ конкатенируемого высказывания: «существует дуга из вершины A в вершину B ». Являясь одномерной формой матрицы смежностей кубит-вектор, ориентирован на решение всех задач дискретного моделирования, синтеза, анализа и оптимизации. Но вопрос заключается в нахождении того класса задач, для которого использование кубитного описания и адресно-ориентированных процедур, исключающих перебор, дает технологические или экономические преимущества.

Синтез (modeling) кубитной модели системы на основе кубитных описаний компонентов

Анализ или моделирование (simulation) кубитной модели структуры на основе заданных входных условий: поиск всех путей от входной или выходной вершины.

Поиск максимального или минимального пути в графе на основе моделирования начальных условий.

Покрывание минимальным количеством путей всех вершин графа для тестирования и верификации цифровых систем.

Поиск дефектов в вычислительных структурах на основе обратного прослеживания некорректной выходной реакции.

Процесс моделирования графа по его кубиту сводится к выполнению единственной процедуры – извлечение состояния из ячейки кубит-вектора по адресу, полученному на основе конкатенации адреса, составленного из кодов вершин истока и стока: $M(A, B) = Q(A * B)$. Такая процедура отвечает на вопрос – существует ли в графе дуга (AB) . Целью процедуры моделирования может быть: 1) нахождение максимального или минимального пути в графе; 2) определение всех путей от входной вершины; 3) вычисление всех вершин приемников для заданного выхода. Если некоторая система содержит совокупность взаимосвязанных графов, тогда процесс моделирования будет использовать следующее выражение: $M_i(A, B) = Q_i(A * B)$.

Интересен также квантовый формат представления графа для эффективного параллельного выполнения элементарных алгебраических операций. Речь идет об алгебро-логических операциях над кубит-векторами одной размерности: конъюнкция (пересечение графов), дизъюнкция (сложение графов), исключающее или (сравнение графов), инверсия (дополнение к графу). Здесь фактически в векторной форме осуществляются манипуляции с дугами на кубит-векторах одинаковой длины. Они могут быть получены, если все графы привести к одному универсальному множеству вершин, которое формирует одинаковое число потенциально возможных дуг. Примеры выполнения

алгебро-логических операций над графами, заданными кубит-векторами, представлены в следующей таблице:

G1	0 1 1 0 0 1 1 1 0 0 0 1 1 0 0 0
G2	0 0 1 1 0 0 0 1 1 1 0 0 1 0 1 0
And(G1,G2)	0 0 1 0 0 0 0 1 0 0 0 0 1 0 0 0
Or(G1,G2)	0 1 1 1 0 1 1 1 1 1 0 1 1 0 1 0
Xor(G1,G2)	0 1 0 1 0 1 1 0 1 1 0 1 0 0 1 0
Not G1	1 0 0 1 1 0 0 0 1 1 1 0 0 1 1 1

Операция хог показывает расстояние между графами, выраженное в количестве единиц, соответствующих кодовому расстоянию по Хэммингу. В данном примере расстояние между двумя графами равно $d(G1, G2) = 8$. Если два или более графа имеют различное число вершин, то для выполнения алгебро-логических операций над ними необходимо коды вершин всех графов привести к одинаковой размерности. Такая процедура создает одинаковое адресное пространство для кодов (A, B) ориентированных дуг всех графов. При этом размерность всех кубит-векторов станет одинаковой, что является необходимым и достаточным условием для алгебро-логических манипуляций над ними. Задача приведения или нормализации всех графовых структур к одинаковой размерности фактически означает перекодировку уже существующих вершин графа в новой системе координат. Если решение задачи нормализации связано с последующим сравнением графов, в целях определения степени их сходства (различия), то необходимо выполнять синтез кодов дуг, приближающий все графы к структуре максимального из них. Данная задача (нормализации, распознавания, кластеризации, классификации графов) чрезвычайно важна для теории, представляет самостоятельный практический интерес и здесь не рассматривается. Тем не менее, ниже представлены четыре квантовых вектора, которые описывают первые четыре графа, приведенные к формату четырех вершин и двухбитовых векторов (пробелы в таблице соответствуют нулевым координатам):

G1	1	1			1	1
G2	1	1	1	1	1	1
G3	1	1	1	1	1	1
G4	1	1	1	1	1	1
And(G3,G1)=G1	1	1			1	1
And(G3,G2)=G2	1	1	1	1	1	1
And(G3,G4)=G4	1	1	1	1	1	1
Xor(G2,G4)=7	1	1	1	1	1	1

Операция логического умножения показывает, что в приведенном формате, состоящем из четырех вершин, все три структуры $(G1, G2, G4)$ являются подмножеством графа $G3$. Кодовое расстояние по Хэммингу между вторым и четвертым графами (последняя строчка таблицы) равно 7 из 16.

Таким образом, представленные результаты имеют определенную научную и практическую значимость: 1) Предложена новая кубитная форма описания графов, которая характеризуется компактностью описа-

ния всех ориентированных дуг, векторным представлением структуры межсоединений, что дает возможность существенно повысить быстродействие методов анализа за счет выполнения параллельных логических операций. 2) Кубитная форма пригодна также для описания логических функциональностей, что делает ее универсальной структурой для применения в проектировании, синтезе и анализе современных специализированных компьютерных систем. 3) Высокая технологичность применения кубитных форм в описании вычислительных устройств дает основания к ее эффективному использованию при решении широкого спектра задач дискретной оптимизации, синтеза и анализа, распознавания и диагностирования, моделирования и тестирования.

Моделирование кубитного описания цифровых структур

Процедура функционального моделирования на Q-векторе логической функциональности сводится к записи в выходную булеву переменную Y состояния бита Q-вектора, адрес которого сформирован на основе конкатенации значений входных переменных: $Y = Q(X) = Q(X_1 * X_2 \dots * X_j \dots * X_k)$. Для моделирования цифровых схем вводится M-вектор состояний линий, как аналог Q-вектора, задающий значения внутренних переменных системы, который связывает Q-векторы логических примитивов (Q-примитивов) в структуру с помощью нумерации входных и выходных переменных каждого функционального элемента [6]. Процедура обработки последнего определяется выражением:

$$M(Y) = Q[M(X)] = Q[M(X_1 * X_2 \dots * X_j \dots * X_k)].$$

С учетом сквозной нумерации Q-примитивов универсальная процедура моделирования текущего i-элемента будет иметь формат

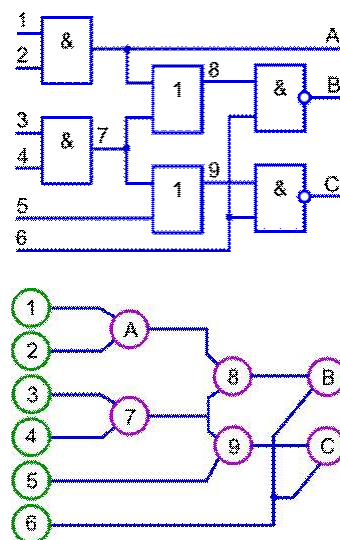
$$M(Y_i) = Q_i[M(X_i)] = Q_i[M(X_{i1} * X_{i2} \dots * X_{ij} \dots * X_{ik_i})]$$

Существенно упрощается алгоритм моделирования цифровой системы, который сводится к процедуре формирования адреса, что дает возможность в $2^n - 1$ раз повысить быстродействие интерпретативного моделирования за счет замены таблиц истинности примитивов на Q-векторы описания только выходных состояний. Кроме того, если логический элемент имеет n входов, то число строк таблицы истинности равно 2^n , что означает – ее размерность равна $d = 2^n \times n$. Учитывая, что длина Q-вектора для любого логического примитива равна 2^n , то выигрыш в объеме памяти для его хранения и обработки составляет:

$$r = \frac{2^n \times n}{2^n} = n.$$

Комбинационная схема (рис. 2) содержит шесть примитивов и три различных логических элемента [6].

Данной схеме соответствует кубит графа и Q-векторы для задания логических примитивов.



$$Q(G) = (.1 \dots 11 \dots 1 \dots 1 \dots \dots \dots),$$

$$G = \{000 - 001, 001 - 010, 001 - 011, 010 - 100, 011 - 101\},$$

$$Q(A) = (0001), Q(7) = (0001), Q(8) = (0111),$$

$$Q(9) = (0111), Q(B) = (1110), Q(C) = (1110).$$

Рис. 2. Граф цифрового устройства

Таким образом, структура графа $Q(G1: A=000, 7=001, 8=010, 9=011, B=100, C=101)$, а также все функциональные элементы $Q(A)-Q(8)$ представлены в форме кубитных покрытий или векторов, что существенно меньше по объему памяти, чем матрица смежностей и таблицы истинности логических элементов.

Выводы

1. Предложенные кубитные модели описания структуры цифровых систем и функций компонентов характеризуются компактностью описания матриц смежностей и таблиц истинности в форме Q-покрытий вследствие унитарного кодирования входных состояний, что позволяет повысить быстродействие программных и аппаратных средств для моделирования вычислительных устройств, благодаря адресной реализации анализа кубитных векторов.
2. Инновационная идея квантовых вычислений заключается в переходе от вычислительных процедур над байт-операндом, определяющим в дискретном пространстве одно решение, к логическим регистровым параллельным процессам над кубит-операндом, одновременно формирующим булеан решений. Это дает возможность определить новые пути создания высокопроизводительных компьютеров параллельного анализа и синтеза структур и компьютерных сервисов.
3. Предложена новая кубитная форма описания графов, которая характеризуется компактностью описания всех ориентированных дуг, векторным представлением структуры межсоединений, что дает возмож-

ность существенно повысить быстродействие методов анализа и моделирования за счет выполнения параллельных логических операций.

4. Высокая технологичность применения квантовых или кубитных форм в описании вычислительных устройств дает основания к ее эффективному использованию при решении широкого спектра задач дискретной оптимизации, синтеза и анализа, распознавания и диагностирования, моделирования и тестирования.

5. Дальнейшие направления исследований связаны с разработкой быстродействующих методов и алгоритмов синтеза, анализа, визуализации, нормализации, распознавания, кластеризации, классификации графово-функциональных структур и систем на основе использования универсальной метрики (например, Манхэттен) приведения и нормализации графов в единой системе координат.

Литература: 1. *Michael A. Nielsen & Isaac L. Chuang*. Quantum Computation and Quantum Information. Cambridge University Press. 2010. 676 p. 2. *Mark G. Whitney*. Practical Fault Tolerance for Quantum Circuits. PhD dissertation. University of California, Berkeley. 2009. 229 p. 3. *Mikio Nishihara*. Quantum Computing. An Overview. Higashi-Osaka: Kinki University, 2010. 53p. 4. *Курош А.Г.* Курс высшей алгебры. М.: Наука. 1968. 426с. 5. *Горбатов В.А.* Основы дискретной математики. М.: Высшая школа, 1986. 311 с. 6. *Hahanov V.I., Litvinova E.I., Chumachenko S.V.* et al. Qubit Model for solving the coverage problem // Proc. of IEEE East-West Design and Test Symposium.

Kharkov. 14-17 September, 2012. P.142-144. 7. *Хаханов В.И., Ваджеб Гариби, Литвинова Е.И., Шкиль А.С.* Кубитные структуры данных вычислительных устройств // Электронное моделирование. 2015. № 1. С.76-99. 8. *Хаханов В.И., Тамер Бани Амер, Чумаченко С.В., Литвинова Е.И.* Кубитные технологии анализа и диагностирования цифровых устройств // Электронное моделирование. 2015. Том 37. № 3 P. 17-40. 9. *Vladimir Hahanov, Wajeb Gharibi, Igor Iemelianov, Dmitry Shcherbin.* «Quantum» Processor for Digital Systems Analysis // Proceedings of IEEE East-West Design & Test Symposium (EWDTs-2015). 2015. Batumi, Georgia. P. 104-110.

Поступила в редколлегию 12.02.2016

Рецензент: д-р техн. наук, проф. Кривуля Г.Ф.

Tamer Bani Amer, аспирант ХНУРЭ. Научные интересы: квантовые вычисления, тестирование и диагностика цифровых систем. Адрес: Украина, 61166, Харьков, пр. Науки, 14, тел. +3805770-21-326.

Чумаченко Светлана Викторовна, д-р техн. наук, профессор, зав. кафедрой АПВТ ХНУРЭ. Научные интересы: математическое моделирование, теория рядов, методы дискретной оптимизации. Адрес: Украина, 61166, Харьков, пр. Науки, 14, тел. + 3805770-21-326, e-mail: ri@kture.kharkov.ua.

Емельянов Игорь Валерьевич, научный сотрудник каф. АПВТ ХНУРЭ. Научные интересы: квантовые вычисления, тестирование и диагностика цифровых систем. Адрес: Украина, 61166, Харьков, пр. Науки, 14, тел. +3805770-21-326.



ИССЛЕДОВАНИЕ КАЧЕСТВА СБОРКИ ПРИЛОЖЕНИЙ ПОД ОПЕРАЦИОННУЮ СИСТЕМУ ANDROID РАЗЛИЧНЫМИ КОМПИЛЯТОРАМИ

НАРОЖНЫЙ В.В., НАРОЖНАЯ Е.В.

Оценивается качество сборки приложения тремя компиляторами: Android Development Tools, Android Studio 2 и Visual Studio 2015. В качестве критериев выбираются быстрдействие и размер полученных приложений. Для чистоты эксперимента описываются три приложения в каждой среде. Приложения реализуют простейший алгоритм поиска простых чисел.

Введение

Android – операционная система (OS Android) для смартфонов, интернет-планшетов, электронных книг, цифровых проигрывателей, наручных часов, игровых приставок, нетбуков, смартбуков, очков Google, телевизоров и других устройств [1].

Классическим средством разработки приложений под OS Android является язык программирования Java [2], принадлежащий компании Oracle.

Рекомендованная Google программная среда разработки приложений на ОС Android до декабря 2014 года была Eclipse [3] Android Development Tools (ADT). На сегодняшний день компания Google активно продвигает новую среду Android Studio 2 (AS2) [4], которая вышла на смену Android Studio с 07.04.2016.

В последнее время появились сообщения о покупке компанией Microsoft компании Xamarin [5]. Xamarin – американская компания в области разработки ПО. Занимается разработкой и поддержкой Mono и инструментов для разработки приложений на языке C# для iOS, Android, Windows, Mac [6].

Компания Xamarin уже несколько лет создавала инструменты разработки под Android (в отличие от рекомендуемого разработчиками Android языка программирования Java, использовался язык программирования C# [7]), которые можно было самостоятельно дополнительно встроить в более ранние версии Visual Studio. В Visual Studio 2015 (VS15) Xamarin включен как один из основных инструментов, что представляет обоснованный интерес в свете покупки компании. Можно предположить, что Microsoft готов свернуть попытки занять рынок мобильных операционных систем для смартфонов и планшетов своей операционной

системой Windows Phone, которая явно не в состоянии отвоевать позиции у Android и iOS. Полное описание исследования приведено на электронном ресурсе [8].

1. Цель исследования

Одной из наиболее сложных задач является оценка качества работы вычислительных устройств. Сложно оценить, какое вычислительное устройство лучше. Классически необходимо выбрать несколько критериев и создать условия, при которых другими критериями можно пренебречь. Тогда проведение исследований устройств по этим критериям можно считать адекватным. Только на основании такого исследования можно делать выводы о качестве. Наличие различных компьютерных архитектур, сред программирования, операционных систем заметно усложняет такую оценку.

Цель и задачи исследования. Цель данного исследования – оценить качество сборки приложения тремя различными компиляторами: ADT, AS2 и VS15.

Для достижения поставленной цели необходимо решить следующие задачи:

- выбрать критерии оценки качества (быстрдействие и размер полученных приложений);
- для чистоты эксперимента написать три приложения на ADT, AS2 и VS15, где реализован простейший алгоритм поиска простых чисел (рис.1);

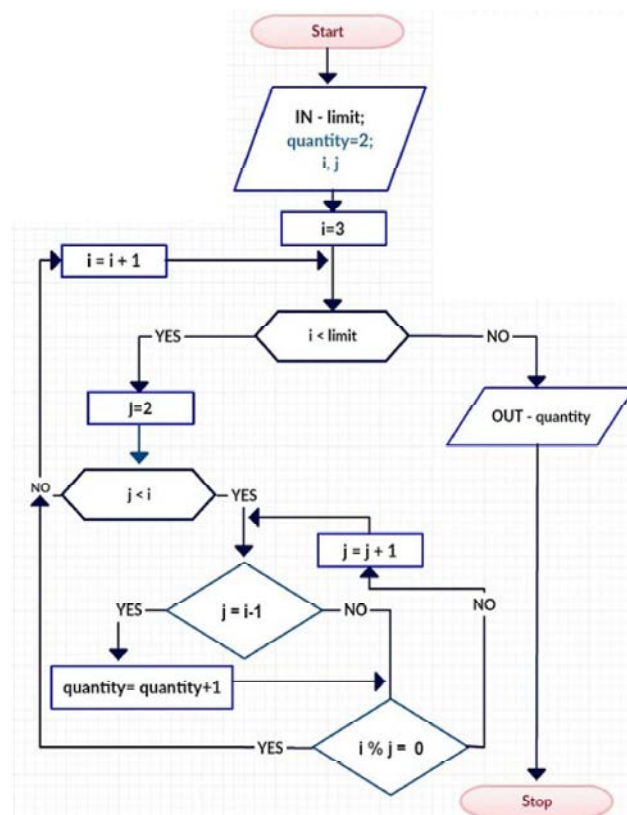


Рис. 1. Алгоритм поиска простых чисел

– в качестве вывода данных использованы время начала работы, время окончания работы и количество (quantity) найденных простых чисел для контроля правильности работы каждого приложения;

– для минимизации возможного воздействия абстракции ООП [9] максимально использовать базовые команды языков Java и C#, а также минимизировано использовать средства ООП;

– для чистоты эксперимента не использованы параллельные потоки, что подвешивает интерфейс на все время поиска.

Таким образом, при поиске относительно большого количества простых чисел (в данных примерах в диапазоне от 3 до limit) можно считать оценку качества сборки приложений адекватной поставленной.

2. Коды программ и интерфейсы пользователей

С учетом всего сказанного были созданы три приложения на ADT, AS2 и VS15. Во всех трех приложениях реализованы одинаковые интерфейсы с использованием файлов данных XML (рис.2) и стилей по умолчанию, характерных для каждой среды.

```

1 <LinearLayout xmlns:android="http://schemas.android
2   android:layout_width="fill_parent"
3   android:layout_height="fill_parent"
4   android:orientation="vertical" >
5   <TextView
6     android:id="@+id/textView0"
7     android:layout_width="match_parent"
8     android:layout_height="wrap_content"
9     android:text="Search prime numbers to " />
10  <EditText
11    android:id="@+id/editText1"
12    android:layout_width="match_parent"
13    android:layout_height="wrap_content"
14    android:text="Input number" />
15  <TextView
16    android:id="@+id/textView1"
17    android:layout_width="match_parent"
18    android:layout_height="wrap_content"
19    android:text="Start time" />
20  <Button
21    android:id="@+id/button1"
22    android:layout_width="match_parent"
23    android:layout_height="wrap_content"
24    android:text="Start" />
25  <TextView
26    android:id="@+id/textView2"
27    android:layout_width="match_parent"
28    android:layout_height="wrap_content"
29    android:text="Stop time" />
30  <TextView
31    android:id="@+id/textView3"
32    android:layout_width="match_parent"
33    android:layout_height="wrap_content"
34    android:text="Quantity prime number" />
35 </LinearLayout>

```

Рис. 2. XML-файл данных ADT

После компиляции все установочные файлы имеют расширение apk (рис.3). Находятся на компьютере в соответствующих папках. Файлы необходимо перенести на устройство для проведения исследований.

Имя	Дата изменения	Тип
findprimeas2-debug.apk	25.04.2016 12:46	Файл "АРК"

Рис. 3. Инсталляционный файл APK

Приложения были написаны с учетом особенностей каждой среды и языков Java (рис.4) и C# (рис.5).

```

package com.example.findprimenumberseclipse;
import ...
public class MainActivity extends Activity
    implements OnClickListener{
    TextView textView1,textView2,textView3;
    Button button;
    EditText editText;
    Calendar c;
    @Override
    protected void onCreate(Bundle savedInstanceState) {
        super.onCreate(savedInstanceState);
        setContentView(R.layout.activity_main);
        button=(Button)findViewById(R.id.button1);
        button.setOnClickListener(this);
        textView1=(TextView)findViewById(R.id.textView1);
        textView2=(TextView)findViewById(R.id.textView2);
        textView3=(TextView)findViewById(R.id.textView3);
        editText=(EditText)findViewById(R.id.editText1);
    }
    @Override
    public void onClick(View v) {
        int searchToTheNumber=Integer.parseInt(editText
            .getText().toString()), quantityPrimeNumber=1;
        c = Calendar.getInstance();
        textView1.setText(""+c.get(Calendar.HOUR)+"":"+
            c.get(Calendar.MINUTE)+" "+c.get(Calendar.SECOND));
        for(int i=3;i<searchToTheNumber;i++){
            for(int j=2;j<i;j++){
                if(j==i-1){quantityPrimeNumber++;}
                if(i%j==0){break;}
            }
        }
        c = Calendar.getInstance();
        textView2.setText(""+c.get(Calendar.HOUR)+" "+
            c.get(Calendar.MINUTE)+" "+c.get(Calendar.SECOND));
        textView3.setText("Quantity prime number: "
            +quantityPrimeNumber);
    }
}

```

Рис.4. Программа, написанная на Java

```

using ...
namespace App1
{
    [Activity(Label = "App1", MainLauncher = true,
        Icon = "@drawable/icon")]
    public class MainActivity : Activity
    {
        TextView textView1, textView2, textView3;
        Button button;
        EditText editText;
        protected override void OnCreate(Bundle bundle)
        {
            base.OnCreate(bundle);
            SetContentView(Resource.Layout.Main);
            button = FindViewById<Button>(Resource.Id.button1);
            textView1=FindViewById<TextView>(Resource.Id.textView1);
            textView2 = FindViewById<TextView>(Resource.Id.textView2);
            textView3 = FindViewById<TextView>(Resource.Id.textView3);
            editText = FindViewById<EditText>(Resource.Id.editText1);
            button.Click += delegate
            {
                int searchToTheNumber = Int32.Parse((
                    (EditText)editText).Text), quantityPrimeNumber = 1;
                textView1.Text = string.Format(DateTime.
                    Now.ToString("HH:mm:ss tt"));
                for (int i = 3; i < searchToTheNumber; i++)
                {
                    for (int j = 2; j < i; j++)
                    {
                        if (j == i - 1) { quantityPrimeNumber++; }
                        if (i % j == 0) { break; }
                    }
                }
                textView2.Text = string.Format(DateTime.
                    Now.ToString("HH:mm:ss tt"));
                textView3.Text = string.Format(
                    "{0} Quantity prime number: ", quantityPrimeNumber);
            };
        }
    }
}

```

Рис.5. Программа, написанная на C#

3. Установка и исследование

Все установочные файлы для OS Android имеют расширение APK. Для установки приложений необходимо зайти в настройки устройства и найти пункт «Разрешить установку из неизвестных источников» (на разных устройствах этот пункт может иметь разное местоположение и название, но, как правило, это в разделе «Безопасность»). Установить галочку на разрешение. Затем загрузить установочные файлы и поочередно сделать на них «тач». Приложения установятся на устройство (рис.6). В данном исследовании использовано устройство Galaxy Ace 3 GT-S7272.



Рис.6. Иконки приложений на устройстве

Теперь можно зайти в «Настройки» - «Диспетчер приложений» и посмотреть размеры, которые занимают приложения в памяти устройства ADT (рис.7), AS2 (рис.8), VS15 (рис.9).

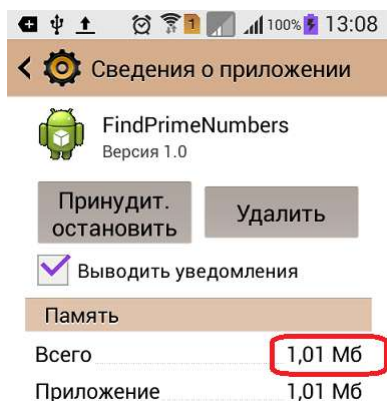


Рис. 7. Размер приложения от компилятора ADT

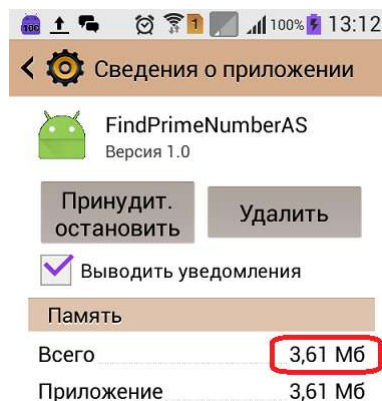


Рис. 8. Размер приложения от компилятора AS2

По требованию издательства цвета, на приведенных скринах, были обращены. Как можно видеть, приложение после установки занимает 1,01 Мб.

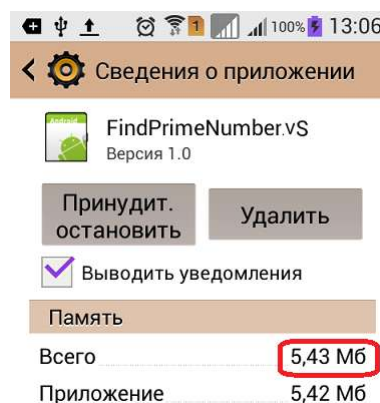


Рис. 9. Размер приложения от компилятора VS15

Итоговый результат можно увидеть на диаграмме (рис.10).

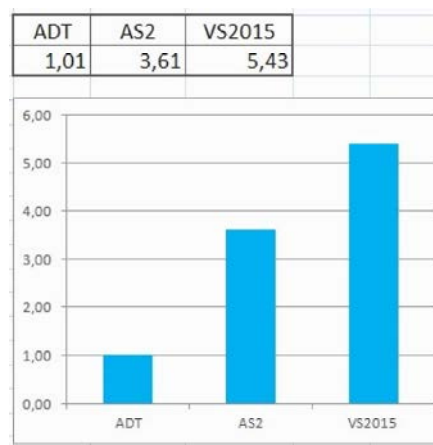


Рис.10. Диаграмма размера установочных файлов установленных приложений

Проверить качество работы компилятора по критерию занимаемого объема внутренней памяти достаточно быстро. Задача по проверке скорости работы приложений занимает значительно больше времени. Для чистоты эксперимента каждый запуск проводился после предварительной очистки памяти. Процесс работы приложений занимал несколько минут (в зависимости от вводимого значения limit). Итоговые результаты приведены на рис. 11.

Выводы

Научная новизна. Предложен метод оценки качества сборки приложения тремя компиляторами: Android Development Tools, Android Studio 2 и Visual Studio 2015. Обоснован выбор критериев (быстродействие и размер полученных приложений) для чистоты эксперимента.

Практическая ценность. Несмотря на огромный опыт компании Microsoft по созданию средств программирования, следует отметить, что по обоим критериям средства разработки приложений на Java лучше. Оба параметра в пользу средств ADT и AS2. В свою

очередь, ADT значительно опережает по размерам установленного приложения. Возможно, AS2 и VS15 можно оптимизировать по этому параметру, но большинство разработчиков этим не занимается, а значит, создаются приложения с явно завышенными объемами включенных ресурсов в конечный продукт. Возможно, это не так заметно для больших приложений, но тут результат очевиден. Также не добавляют оптимизма процесс установки VS15 или AS2 (явно сложнее процесса установки ADT) и значительно более высокие требования к ресурсам компьютера.

limit	ADT,сек	AS2,сек	VS15,сек
60000	9	9	10
80000	16	16	18
100000	25	25	27
120000	34	34	37
140000	47	47	51
160000	61	61	66
180000	77	77	83
200000	94	94	102
220000	112	112	122

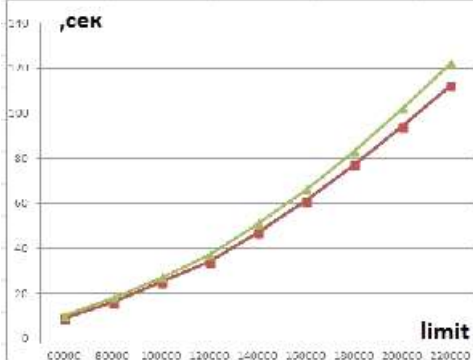


Рис. 11. Показатели скорости работы приложений

Учитывая все сказанное выше, можно сделать вывод, что явно прослеживается очередной виток от качества сборки приложения к увеличению числа различных фишек в средах разработки.

Литература: 1.*Android*: электронный ресурс. Точка доступа https://ru.wikipedia.org/wiki/Android_2. 2.*Java*: электронный ресурс. Точка доступа https://uk.wikipedia.org/wiki/Java_3. 3.*Eclipse*: электронный ресурс. Точка доступа https://uk.wikipedia.org/wiki/Eclipse_4. 4.*Android Studio*: электронный ресурс. Точка доступа https://ru.wikipedia.org/wiki/Android_Studio. 5.*Microsoft to acquire Xamarin and empower more developers to build apps on any device*: электронный ресурс. Точка доступа <http://blogs.microsoft.com/blog/2016/02/24/microsoft-to-acquire-xamarin-and-empower-more-developers-to-build-apps-on-any-device/>. 6.*Xamarin*: электронный ресурс. Точка доступа https://ru.wikipedia.org/wiki/Xamarin_7. 7.*C#*: электронный ресурс. Точка доступа https://ru.wikipedia.org/wiki/C_Sharp_8. 8.*Полное описание исследований*: электронный ресурс <http://androidresearchua.blogspot.com/>. 9.*Абстрактный тип данных*: электронный ресурс. Точка доступа https://ru.wikipedia.org/wiki/Абстрактный_тип_данных

Поступила в редколлегию 16.02.2016

Рецензент: д-р техн. наук, проф. Моисеенко В.И.

Нарожный Виталий Васильевич, канд. техн. наук, доцент кафедры Информационных технологий и мехатроники Харьковского национального автомобильно-дорожного университета. Научные интересы: исследование особенностей операционной системы Android. Адрес: Украина, 61000, Харьков, ул. Ярослава Мудрого, 25, тел.:+380508577597.

Нарожная Екатерина Витальевна, студентка кафедры Информационных технологий и мехатроники Харьковского национального автомобильно-дорожного университета. Научные интересы: исследование особенностей операционной системы Android. Адрес: Украина, 61000, Харьков, ул. Ярослава Мудрого, 25, тел.:+380667487704.

ФОРМАЛИЗАЦИЯ НЕЯВНОГО ПРОЦЕДУРНОГО ЗНАНИЯ

ЛЕВЫКИН В.М., ЧАЛАЯ О.В.

Рассматриваются особенности неявного процедурного знания. Предлагается модель такого знания, содержащая явную составляющую в форме последовательностей действий и неявную составляющую в форме неформализованных отношений между сущностями предметной области, которые задают необходимое и достаточное условие для выполнения таких действий. Предлагается подход к формализации неявных процедурных знаний, основанный на поэтапном дополнении явного представления формализуемым неявным контекстом предметной области.

Введение

В настоящее время при построении систем управления знаниями большое внимание уделяется решению проблемы формализации и использованию неявных знаний, присущих естественному интеллекту (ЕИ). Неявное знание рассматривается как знание, которое неотделимо от человека, существует, передается и используется в неформализованной форме [1-4]. Для того чтобы такое знание можно было использовать в системах искусственного интеллекта, необходимо выполнить его экстернализацию (преобразование в явную форму). Для выполнения такого преобразования необходимо системно описать неявное знание, выделив его признаки, элементы, а также взаимосвязи между элементами. Последнее предполагает построение общей модели неявного знания, а также моделей его составляющих.

Исследование свойств и структуры неявных знаний, выполненное в работах [4-7], показывает, что такие знания содержат как скрытые зависимости между сущностями предметной области, так и наборы действий над указанными сущностями, выполняемые с учетом неявных зависимостей. В то же время вопросам формализации таких процедурных знаний с учетом неявного контекста предметной области не уделяется должного внимания.

Следовательно, проблематика экстернализации неявных знаний в целом и неявных процедурных знаний в частности является актуальной.

Постановка задачи

Экстернализация неявного процедурного знания основана на использовании скрытого контекста при выполнении последовательностей действий в рамках процедур. Для использования такого неявного знания в системах искусственного интеллекта необходимо структурировать контекст, выделив его явную и формализовав неявную составляющую, а также определив связь этих составляющих с возможными в данном контексте последовательностями действий.

Целью данного исследования является построение модели неявного процедурного знания, позволяющей формализовать возможные и допустимые действия в условиях неявного контекста, который содержит неформализованные взаимосвязи между сущностями предметной области.

Неявное процедурное знание

Неявное процедурное знание объединяет множество процедур решения прикладных задач, представляемых в виде одной или нескольких последовательностей действий и учитывающих неформализованные возможности и ограничения в заданной предметной области. Данный вид знания соответствует навыкам в ЕИ [8] и представляет собой набор возможных и допустимых действий, а также последовательностей действий в заданном частично формализованном контексте.

Под контекстом в данном случае будем понимать систему зависимостей между свойствами сущностей предметной области, которая определяет возможность и допустимость используемых при решении прикладных задач наборов действий (процедур).

Выявление неявных процедурных знаний позволяет решить многие прикладные задачи, в частности в области управления производством.

Например, транснациональная фирма «Volvo IT» занимается поиском неявных контекстных зависимостей, определяющих различную эффективность одних и тех же бизнес-процессов в подразделениях, находящихся в разных странах. Во всех подразделениях реализованы однотипные бизнес-процессы, однако время обслуживания клиентов значительно отличается. Сведения о выполняемых последовательностях действий формализованы и фиксируются в формате файлов логов. Существующие методы process mining позволяют построить модели таких бизнес-процессов. Однако для выявления причин отличий между процессами необходимы дополнительные неформализованные знания, которые не отражены в существующих моделях [9].

Процедурные неявные знания характеризуются следующими особенностями:

- допустимые действия, составляющие процедуры решения прикладных задач, представляются в формальном виде;
- выбор подходящих процедур в конкретной ситуации осуществляется на основе неформализованных зависимостей;
- в качестве исходных данных при выборе процедур решения прикладных задач используются как явные, так и неявные реляционные знания.

Практическая ценность данного вида знаний состоит в том, что использование неявного контекста позволяет выделить более эффективные по критериям временных и материальных затрат последовательности действий, используемые при решении практических задач.

Контекст, в котором реализуются процедуры, может двояко влиять на выбор возможных последовательностей действий.

Во-первых, неформализованные контекстные взаимосвязи могут задавать дополнительные ограничения на реализацию операций в конкретных условиях предметной области. Семантика таких ограничений может быть различной. В частности, за счет неявных ограничений могут отсекаются типовые, но недостаточно эффективные последовательности действий или же последовательности, затрудняющие работу исполнителя. В данном случае прослеживается аналогия между работой новичка, использующего общепринятые способы решения задач, и эксперта, эффективно решающего задачи за минимальное время.

Во-вторых, такие контекстные зависимости могут создавать условия для построения более эффективных процедур решения существующих задач. Указанные процедуры не могут быть получены на основе использования только явных знаний, поскольку последние не полностью описывают предметную область. Новые способы применения определяются скрытыми ранее свойствами сущностей предметной области, т.е. из экстернализованных знаний о наборе свойств сущностей и базовых знаний о предметной области могут быть выведены процедуры использования таких сущностей.

Проиллюстрируем данный вид неявных знаний на примере экстернализации знаний в естественном интеллекте. Для ЕИ рассмотренные ограничения и условия обычно связаны с физическими и ментальными характеристиками людей исполнителей, а также физическими свойствами сущностей предметной области. Пример отклонения от обычной последовательности решения задачи: обмерзание колонны, в которой происходит химическая реакция. В управляющую программу способы устранения возникшей ситуации не заложены. Очевидные неявные знания: обмерзание колонны на определенной высоте. Дополнительные действия, выполняемые после выявления этих знаний: устранение обмерзания. Исходные знания являются очевидными реляционными неявными, поскольку колонна доступна и можно измерить температуру по всей ее длине, установив зону обмерзания. Однако на практике такие измерения не проводились при создании основанной на правилах системы управления и потому не были реализованы в рамках управляющего программного обеспечения. Исполнитель же обладает неявными знаниями, о новых способах решения задачи в конкретной ситуации.

Таким образом, неявные процедурные знания K^P включают в себя:

- наборы процедур, учитывающих неявные ограничения на допустимые последовательности действий в текущем состоянии предметной области;
- наборы процедур, использующих неявные условия, представленные в форме скрытых отношений между сущностями предметной области.

Процедурные знания первого вида K_L^P задают набор альтернативных i -процедур (последовательностей действий) D_{ij} , допустимых в рамках заданного контекста C_j :

$$K_L^P = \bigvee_i (D_{ij} | C_j). \quad (1)$$

Последовательность действий D_{ij} с учетом темпоральной составляющей неявных знаний может быть формально представлена средствами модальной темпоральной логики, например LTL, CTL.

Процедурные знания второго вида K_C^P задают набор дополнительных недокументированных ранее возможностей выполнения i -процедур в j -контексте с учетом неявной составляющей контекста:

$$K_C^P = \bigwedge_{i=1}^I (D_{ij}^o | \bigwedge_{n=1}^N k_{n,j}^R), \quad (2)$$

где $k_{n,j}^R$ – очевидное или скрытое неявное отношение между сущностями предметной области, истинность которого позволяет реализовать I новых последовательностей действий D_{ij}^o в контексте C_j .

Контекст C_j содержит как явную, так и неявную составляющие:

$$C_j = (\bigwedge_n k_{n,j}^R) \wedge (\bigwedge_s k_{s,j}^E) \quad (3)$$

$$\forall k_{n,j}^R \in K_{IC}^R, \forall k_{s,j}^E \in K^E \exists O_j \subseteq O,$$

где O – набор сущностей (например, объектов) предметной области, представляемых в виде набора атрибутов a_n : $O = \{o_y\}, o_y = \{a_n\}$; $k_{n,j}^R, k_{s,j}^E$ – соответственно неявные и явные зависимости между сущностями предметной области, истинность которых зависит от значений переменных, соответствующих атрибутам объектов.

Как видно формального определения C_j , ограничение истинности контекста в выражении (1) будет включать в себя дополнительные зависимости после формализации неявных реляционных знаний. Тогда, с учетом приведенного выше описания очевидных и скрытых разновидностей реляционных неявных знаний, процедурное неявное знание K^P определяется через набор допустимых и возмож-

ных последовательностей действий следующим образом:

$$K^P = K_C^P \wedge K_L^P = \bigvee_i (D_{ij} \wedge (\bigwedge_n k_{n,j}^R) \wedge (\bigwedge_s k_{s,i}^E)) \wedge (\bigwedge_{i=1}^I (D_{ij} \wedge (\bigwedge_{n=i}^N k_{n,j}^R))), \quad (4)$$

$$\exists O_j \subseteq O.$$

В неявных процедурных знаниях процедура выбора последовательности действий обычно не выделена в силу ее неформального характера. Поэтому данная процедура может быть представлена через выделенные и формализованные ограничения на допустимые последовательности действий. Указанные ограничения отражают связь между свойствами сущностей предметной области, а также возможными значениями этих свойств. Следовательно, указанные зависимости изменяются во времени. Это означает, что они выражаются через конкретные начальные условия, отражающие значения свойств сущностей предметной области с привязкой к конкретным моментам времени.

Проведенная структуризация и формализация неявного процедурного знания позволяет сформировать общий подход к его выявлению.

Для обоснования предлагаемого подхода рассмотрим особенности неявных знаний как элемента естественного интеллекта.

В общем случае неявное знание может быть открыто заново на основе изучения формализованной части общего знания. Однако, как уже отмечалось ранее, неявное знание имеет форму, отличную от символической, используемой при описании явного знания. Поэтому неявное знание обычно передается напрямую, без промежуточного формального описания [1,2]. При этом могут передаваться знания как о наборе взаимосвязей между сущностями предметной области, так и процедурные знания о действиях, которые могут быть выполнены с этими сущностями.

Поскольку при исследованиях в области естественного интеллекта совокупность таких действий обычно рассматривается как навык, то при формализации неявного процедурного знания целесообразно использовать выделенные схемы передачи навыков [8].

Обобщенная схема приобретения навыков, предложенная в работе [10], содержит основные шаги приобретения человеком знаний о последовательности действий, применимых при решении практических задач в заданной предметной области. В основе данной схемы лежит идея о постепенном дополнении явных знаний неявными путем их передачи от носителя к носителю.

Предлагаемый подход к формализации неявного процедурного знания основан на использовании

совокупности явных и формализованных реляционных неявных знаний. Набор явных знаний дополняется путем формализации неявных отношений между сущностями предметной области. Последние составляют формализованный неявный контекст. Получившаяся система знаний дополняется формализованной неявной процедурной составляющей, отражающей последовательность решения задач с учетом имеющихся навыков в естественном интеллекте. Данный подход включает в себя следующие этапы.

Этап 1. Систематизация явных формализованных правил, задающих допустимые последовательности действий при решении поставленных практических задач. Под систематизацией в данном случае будем понимать отбор множества правил, пригодных для решения поставленных задач в заданной предметной области, а также последующую проверку таких правил на полноту и непротиворечивость. На данном этапе определяются знания, необходимые для решения задач в типовых ситуациях стандартным (наиболее часто используемым) способом.

Этап 2. Дополнение сформированной на предыдущем этапе системы правил экстернализованными неявными зависимостями, отражающими свойства сущностей предметной области, а также взаимосвязи между указанными свойствами. Такие свойства и взаимосвязи являются элементом реляционных неявных знаний. Указанные зависимости задают дополнительные ограничения и способы применения, изменяющие последовательность использования уже известных правил, полученных на первом этапе. Очевидно, что неявные реляционные зависимости должны быть формализованы на данном этапе.

Этап 3. Выделение неявного контекста для подмножеств явных правил. Контекст (3) представляет собой модель предметной области, дополненную неявными связями между сущностями. Экстернализация неявной составляющей контекста позволяет в общем представлении знаний о предметной области сформировать новые, ранее неизвестные ситуации. Это даст возможность использовать ситуационный подход и формализовать совокупное представление знаний в виде иерархии ситуаций. Рассматриваемые неявные процедурные знания включают условия выполнения действий, которые могут быть истинны лишь в определенных моменты времени. Иными словами, такие знания имеют темпоральную составляющую и потому предлагаемая иерархия ситуаций формируется как по темпоральному, так и по объектным признакам.

Этап 4. Формирование общего контекста для рассматриваемой предметной области путем дополнения представления знаний неявными правилами использования реляционных знаний. Применительно к естественному интеллекту указанные правила позволяют выбрать контекст для исполнительской с

различной квалификацией, поскольку последние обладают отличающимися наборами навыков. Выбор контекста определяет дальнейший сценарий поведения при решении практических задач.

Этап 5. Формирование сценариев реализации практических задач для различного контекста. На данном этапе сценарий формируется в виде последовательности ситуаций, связанных с конкретным контекстом. Различные сценарии отражают варианты решения практических задач при отличающихся подмножествах ограничений.

В целом предлагаемый подход к построению представления знаний о предметной области на основе объединения явных и неявных процедурных знаний направлен на разделение последовательности действий, позволяющих эффективно решить практическую задачу, и правил, обосновывающих эффективность такой последовательности в конкретных ситуациях. Обоснование и формализация связи между действиями и ситуацией выполняется однократно. Последовательность действий в конкретной ситуации может быть использована много раз при решении практических задач.

Иными словами, эффективная контекстно – зависимая последовательность может быть получена путем логического вывода на явных и формализованных неявных реляционных знаниях и в дальнейшем многократно использована при решении текущей задачи.

Выводы

Предложена модель неявного процедурного знания, содержащая явную составляющую в форме последовательностей действий и неявную составляющую в форме неформализованных отношений между сущностями предметной области, которые задают необходимое и достаточное условия для выполнения этих действий. Предложен подход к формализации неявных процедурных знаний, основанный на поэтапном дополнении явного представления формализуемыми неявными зависимостями, общим контекстом предметной области, а также типовыми сценариями решения прикладных задач в заданном контексте.

Полученные модель и подход служат теоретической основой для построения контекстно-зависимого представления знаний, позволяющего выполнить экстернализацию неявных знаний и, тем самым, расширить возможности систем искусственного интеллекта.

Литература: 1. *Polanyi M. Personal Knowledge: Towards a Post-Critical Philosophy/ M. Polanyi. University of Chicago Press, 1958. 493 p.* 2. *Polanyi M. The Tacit Dimension / M. Polanyi. – University of Chicago Press, 1966. 104 p.* 3. *Нонака И. Компания - создатель знания. Зарождение и развитие инноваций в японских фирмах / И. Нонака, Х. Такеучи. М., 2003. 384 с.* 4. *McInerney C. “Knowledge Management and the Dynamic Nature of Knowledge” / Claire McInerney // Journal of the American Society for Information Science and Technology. 2002. № 53 (12) . 1009–1018.* 5. *Smith E. A. The role of tacit and explicit knowledge in the workplace / Elizabeth A. Smith // Journal of Knowledge Management. 2001. №5(4). P. 311–321.* 6. *Gascoigne N. and Thornton T. Tacit knowledge / Neil Gascoigne and Tim Thornton. Acumen. 2013. 207 p.* 7. *Collins H. Tacit and Explicit Knowledge / Harry Collins. The University of Chicago Press. 2010. 186 p.* 8. *Guranzon B. Knowledge, Skill and Artificial Intelligence/ B. Guranzon, I. Josefson //The Springer Series on Foundations and Applications of Artificial Intelligence, 1988. 193 p.* 9. *BPI Challenge 2013, open problems. Ghent University. Dataset. [Электронный ресурс] / Режим доступа: <http://dx.doi.org/10.4121/uuid:3537c19d-6c64-4b1d-815d-915ab0e479da>* 10. *Dreyfus H. Mind Over Machine: The Power of Human Intuition and Expertise in the Era of the Computer / H. Dreyfus, S. Dreyfus// New York: Free Press, 1986. p. 50.*

Поступила в редколлегию 12.02.2016

Рецензент: д-р техн. наук, проф. Филатов В. О.

Левыкин Виктор Макарович, зав. кафедрой информационно-управляющих систем ХНУРЭ. Адрес: Украина, 61166, пр. Науки, 14, тел. (057) 70-21-451.

Чалая Оксана Викторовна, канд. эконом. наук, доцент кафедры информационно-управляющих систем ХНУРЭ. Адрес: Украина, 61166, пр. Науки, 14, тел. (057) 70-21-451.

МЕТОДОЛОГИЯ ОЦЕНКИ ВЛИЯНИЯ КИБЕРАТАК НА БЕЗОПАСНОСТЬ ВИДЕОИНФОРМАЦИОННОГО РЕСУРСА В ТЕЛЕКОММУНИКАЦИОННЫХ СИСТЕМАХ

БАРАННИК В.В., ПОДЛЕСНЫЙ С.А.,
ШУЛЬГИН С.С.

Излагаются основные этапы методологии оценки влияния кибератак на безопасность видеоинформационного ресурса. Анализируются современные угрозы информационно-телекоммуникационных систем. Обосновывается актуальность защиты видеоинформационного ресурса. Описывается проявление наиболее существенных атак типа “Distributed Denial of Service”. Предлагаются основные требования для технологии кодирования, которые обеспечивают требуемую защиту.

1. Введение

Стремительное развитие информационных технологий привело к широкому распространению передачи мультимедийных данных по информационно-телекоммуникационным сетям. Применение мультимедиа для передачи данных позволяет увеличить качество восприятия информации, так как человек уделяет наибольшее внимание визуальной информации. В настоящее время в интересах ведомственных организаций и профильных министерств передача мультимедийных файлов в одном направлении используется в системах видеомониторинга, а двунаправленная передача применяется в системах видеоконференцсвязи. На сегодняшний момент данные системы очень широко применяются в государственных ведомственных структурах для повышения качества управления по соответствующим направлениям деятельности. Как показывает практика в обычных условиях, особенно в кризисных ситуациях, повышаются угрозы применения злоумышленниками кибератак. В то же время вопросы, которые касаются исследований актуальности и значимости влияния различных кибератак на характеристики целостности и оперативности видеоинформационного ресурса в телекоммуникационной системе, недостаточно рассмотрены.

Поэтому необходимо разработать методологию, которая учитывает существующие механизмы функционирования телекоммуникационной системы и действия кибернетических атак на видеоинформационный ресурс.

Цель исследования состоит в том, чтобы предложить пути развития методов кодирования для обеспечения требуемой защиты информационной безопасности.

2. Описание процесса обработки видеопотока и доставки видеоинформационного ресурса в телекоммуникационной системе

Процесс обработки видеопотока и доставки данных в телекоммуникационной системе состоит из следующих основных этапов (рис. 1).

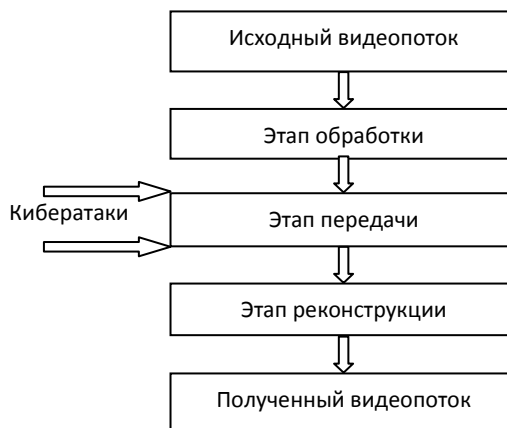


Рис. 1. Этапы обработки видеопотока

Рассмотрим каждый этап данного процесса.

1. *Этап обработки.* На данном этапе происходит подготовка изображения для его передачи в телекоммуникационной системе так, чтобы обеспечить необходимое качество на приемной стороне.

Оператор обработки функционирует на основе нескольких механизмов. В процессе сжатия видеоизображения используются следующие механизмы: устранение пространственной избыточности изображения, устранение временной избыточности видео, использование более низкого цветового разрешения при цветоразностном представлении изображения, повышение информационной плотности результирующего цифрового потока. Основными выходными параметрами на этапе обработки исходного видеопотока являются:

BR – битовая скорость кодированных данных;

Φ_K – время кодирования;

PSNR – отношение пикового уровня сигнала к шуму для изображения.

2. *Этап передачи.* На данном этапе производится передача видеопотока от источника к потребителю с применением телекоммуникационных технологий.

При использовании телекоммуникационных технологий учитываются следующие параметры: C – пропускная способность; P – вероятность потери бита при передаче данных.

Процесс функционирования информационно-телекоммуникационных систем определяется следующими операторами: $F_{AQM}(\lambda, \mu, W, S, QoS)$ – оператор обработки очередей пакетов; $F_{CRC}(P)$ – оператор проверки пакетов на ошибки при проверке пакетов; $F_{ROUTE}(Proto, AddrDst, PortDst, AddrSrc, PortSrc, TTL)$ – оператор маршрутизации пакетов.

Оператор обработки очередей пакетов функционирует на основе нескольких механизмов. В классическую схему системы управления трафиком входят следующие механизмы: ограничение исходящего трафика (shaping), планирование (scheduling), классификация (classifying), ограничение входящего трафика (policing), уничтожение (dropping), маркирование (marking) пакета. Данный оператор зависит от следующих характеристик:

λ – интенсивность входного потока;

μ – интенсивность обслуживания выходного потока;

W – емкость буфера коммутационного устройства;

S – размер пакета;

QoS – класс обслуживания.

Оператор маршрутизации пакетов функционирует на основании полученной информации об адресации сетевого уровня, которую он извлек из пакета данных. Этот оператор зависит от следующих характеристик:

Proto – наименование протокола;

AdrDst, PortDst – адрес и порт назначения;

AdrSrc, PortSrc – адрес и порт источника;

TTL – время жизни пакета.

3. *Этап реконструкции.* На данном этапе происходит декодирование информации и вывод изображения пользователю.

Оператор реконструкции функционирует на основе нескольких механизмов. В процессе реконструкции видеоизображения используются следующие механизмы: получение декодированного остаточного макроблока в процессе деквантования и обратного преобразования; получение восстановленного макроблока для отображения в декодированном кадре.

Для оценки качества воспроизведения видео при декодировании существенными являются следующие параметры:

$T_{3П}$ – задержка пакетов;

δ – джиттер пакетов;

$v_{ПП}$ – количество потерянных пакетов.

Согласно рекомендациям международного союза электросвязи (МСЭ) [1] для системы видеосвязи задержка пакета не должна превышать $T_{3П} \leq 150 \text{ мсек}$, для видеопотока задержка пакета не должна превышать $T_{3П} \leq 10 \text{ сек}$, джиттер пакетов не должен превышать $\delta \leq 80 \text{ мсек}$, количество потерянных пакетов не должно превышать $v_{ПП} \leq 1\%$.

Эти параметры зависят от динамической характеристики среды передачи данных и операторов информационно-телекоммуникационных систем и даже при нормальных условиях (при отсутствии какого-либо вмешательства в работу системы извне) функционирования телекоммуникационных устройств возможны потери видеoinформационного ресурса. К приме-

ру, вероятность потери бита информации при передаче в нормальных условиях в проводных системах составляет около 10^{-8} , а в беспроводных – около 10^{-3} [2].

Теперь рассмотрим, как изменятся параметры оценки качества воспроизведения видео в случае применения кибератак.

3. Влияние DDoS атак на информационно-телекоммуникационную систему

На информационно-телекоммуникационную систему могут воздействовать злоумышленники. Целью кибернетических атак злоумышленников является нарушение таких свойств информационной безопасности как целостность и доступность видеoinформационного ресурса.

В течение 2014 года специализированным структурным подразделением Государственного центра защиты информационно-телекоммуникационных систем (ГЦЗ ИТС) Государственной службы специальной связи и защиты информации Украины (Госспецсвязи) CERT-UA были приняты меры по реагированию на 216 компьютерных инцидентов [3]. Статистика по типам угроз и секторам возникновения приведена в табл. 1, 2. Данные приводятся в отношении тех инцидентов, по поводу которых CERT-UA было сообщено в установленном порядке.

Таблица 1
Соотношение угроз

Типы угроз	Количество	Доля %
DDoS	51	24
Несанкционированный доступ	39	18
Фишинг	30	14
Malware	25	12
Advanced Persistent Threat (APT)	25	12
Другое	46	21
Итого	216	100

Таблица 2
Распределение угроз по секторам

Типы угроз	Принадлежность сектора				
	UAG OV	UAC OM	FGO V	FCO M	UAC TZ
DDoS	43	2	3	3	0
Несанкционированный доступ	33	3	3	0	0
Фишинг	0	6	1	23	0
Malware	7	10	0	1	7
Advanced Persistent Threat (APT)	21	3	1	0	0
Бот-сети	5	6	2	2	1
Уязвимости	13	1	1	0	2
Мошенничество	2	6	0	0	0
Утечка информации	0	3	0	0	0
Другое	0	2	0	0	0
Итого	124	42	11	29	10

Как видно из указанных таблиц, наиболее распространенными видами кибератак (43 для украинского государственного сектора, 2 для украинского коммерческого сектора, 3 для зарубежного государственного

сектора, 3 для зарубежного коммерческого сектора) являются атаки типа DDoS-атака. Это связано, прежде всего, с легкостью реализации данного типа атаки.

При описании модели злоумышленника считается, что его тип будет профессионал или сотрудник предприятия. Для профессионала характерна способность добывать сведения об информационно-телекоммуникационной системе, планировать и готовить вторжение, у него имеется специальный набор средств для осуществления кибератак. Для сотрудника предприятия характерна способность добывать сведения об информационно-телекоммуникационной системе, планировать и готовить вторжение, произвести саботаж работы информационно-телекоммуникационной системы. У него имеется набор самодельных или доступных средств для осуществления кибератак ТСО, он действует скрытно в рабочее время.

В условиях применения атаки на телекоммуникационное устройство дополнительно действует оператор влияния атаки $F_A(\lambda_A, S_A, Proto_A)$.

Оператор влияния атаки на телекоммуникационное устройство функционирует на основании следующих характеристик:

λ_A – интенсивность атаки (количество пакетов),

S_A – размер пакета,

$Proto_A$ – наименование протокола.

На данный момент существует достаточно много различных видов атак на отказ, каждая из которых использует определенную особенность построения сети или уязвимости программного обеспечения.

Эти атаки могут осуществляться путем непосредственной пересылки большого количества пакетов (UDP, ICMP flood), использование атак на промежуточные узлы (Smurf, Fraggle), передачи слишком длинных пакетов (Ping of Death), некорректных пакетов (Land) или большого количества трудоемких запросов [4]. Заметим, что в последнее время происходит развитие этого направления деятельности и появление новых видов и способов атак. Из последних тенденций можно отметить появление атак ухудшения качества (Quality Reduction Attack) и низкочастотных атак (Low Rated Attack) Поэтому необходимы исследования и разработки новых методов противодействия.

Основные существующие классы атак достаточно хорошо изучены. Атаки классифицированы согласно протоколам, по которым они осуществляются. Выделены следующие атаки: SYN flood, TCP reset, ICMP flood, UDP flood, DNS request, CGI request, Mail bomb, ARP storm и атаки на алгоритмическую сложность.

Истощение ресурсов сети заключается в пересылке большого количества пакетов в сеть жертвы. Они уменьшают ее пропускную способность для законных пользователей. Существует несколько видов таких атак:

UDP/ICMP flood заключается в пересылке значительного количества крупных (фрагментированных) пакетов по протоколам UDP/ICMP;

Smurf/Fraggle заключается в пересылке пакетов UDP/ICMP ECHO на широкий диапазон адресов со сфальсифицированного IP адреса. При этом на адрес жертвы приходит большое количество пакетов-ответов.

Истощение ресурсов узла заключается в пересылке трудоемких или некорректных запросов жертве. К этому виду относятся следующие атаки:

TCP SYN – сознательное прерывание процесса установления соединения и создание большого количества полуоткрытых TCP/IP соединений (поскольку это число ограничено, то узел перестает принимать запросы на соединение);

Land – пересылка пакета TCP SYN с одинаковыми адресами получателя и источника и портами (при посылке таких пакетов узел с Windows NT зависает);

Ping of Death – посылка пакета “ping” очень большой длины, который ОС не может обработать;

пересылка некорректных пакетов, при обработке которых на узле могут возникнуть ошибки;

пересылка трудоемких запросов для загрузки узла.

Достаточно эффективной формой отраженной атаки является использование серверов доменных имен (Domain Name System (DNS) servers). Схематически данная атака показана на рис. 2.

Эти серверы нужны для хранения и предоставления по требованию различных записей (Resource records (RR)) с именами доменов Интернет. В такой записи могут быть данные типа TXT RR, в которые администратор может внести произвольный текст, типа A RR, где определяется отображение имени в 32-битную IP адрес, и типа SOA1 RR, где определяется имя домена Интернет и другая сопутствующая информация. Важнейшей функцией DNS серверов является трансляция доменных имен в IP адреса. После получения запроса на IP адрес DNS сервер пытается найти соответствующую запись в своей базе данных. Если это не удалось, то запрос рассылается по всем известным ему DNS серверам. Эта процедура называется рекурсивным запросом. Размеры запроса и ответа могут существенно отличаться. Обычно ответ содержит первоначальный запрос и ответ. Это означает, что ответ всегда больше запроса. Более того, ответ может содержать различные типы RR данных и некоторые из них могут иметь значительный объем. Теоретически, начальный трафик мощностью 140 Mb/s с ботнет может привести к потоку DNS ответов мощностью 10 Gb/s.

Как видно из указанного, при существующей модели обработки и доставки видеoinформационного ресурса телекоммуникационное устройство будет обрабатывать большой поток информации. При существующих механизмах его работы происходит задержка переда-

чи пакетов с возможным их отбрасыванием. Основным результатом DDoS-атаки на телекоммуникационное устройство заключается в изменении времени прохождения и общего количества пакетов информации для видеoinформационного ресурса.

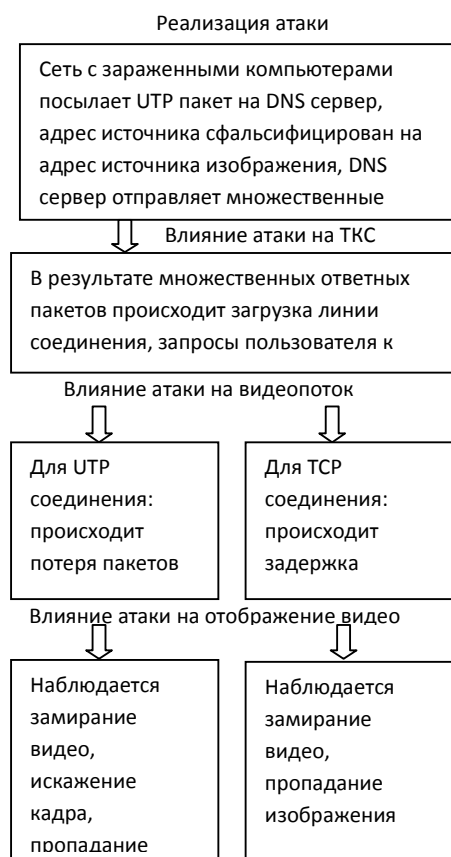


Рис. 2. Влияние атаки типа “DNS Amplification Attacks” на передачу видеoinформационного ресурса в информационно-телекоммуникационной системе

Для предупреждения возникновения атаки типа DDoS-атака могут применяться системы обнаружения и предотвращения вторжений, управляемые коммутаторы со списками контроля доступа и резервирования линий связи между отдельными узлами [5].

К недостаткам перечисленных методов относятся вносимые через обработку пакетов задержки, и как результат ограничение скорости передачи сигнала в сети, увеличенная стоимость оборудования и подписки на сигнатуры, данными методами не обеспечивается защита от замаскированных вирусных атак.

Поэтому необходимо разработать новый метод прогнозирования и локализации атак, который учитывает этап кодирования источника и обеспечивает требуемую степень соответствия полученной информации показателям класса обслуживания в условиях применения DDoS-атак, когда возможны задержки передачи пакетов с возможным их отбрасыванием.

4. Выводы

1. Разработана методология оценки влияния кибератак на безопасность видеoinформационного ресурса в телекоммуникационной системе.

Эта методология основана на описании процесса обработки и доставки данных, учитывает существующие операторы и механизмы процесса. Она призывает разработать новый метод прогнозирования и локализации атак, который учитывает этап кодирования источника и обеспечивает требуемую степень соответствия полученной информации показателям класса обслуживания.

2. Обосновано, что DDoS-атаки влияют на передачу в информационно-телекоммуникационной системе. Это приводит к потере целостности и доступности видеoinформационного ресурса.

3. Показано, что существующие технологии противодействия DDoS-атак не обеспечивают в полной мере их локализацию и предупреждение (т.е. они работают с последствием, запаздыванием, после распознавания факта применения атаки). Также немаловажен тот факт, что в случае выявления кибератаки существующие системы противодействия DDoS-атак отсекают входящий поток данных, из-за чего происходит потеря “полезной” информации, т.е. ухудшаются такие характеристики безопасности видеoinформационного ресурса, как целостность и доступность.

Литература: 1. Рекомендации Международного союза электросвязи ИТУ-Т G.1010 “End-User multimedia QoS categories”. 2. Олифер В.Г. Компьютерные сети. Принципы, технологии, протоколы: Учебник для вузов / В.Г. Олифер, Н.А. Олифер. СПб.: Питер, 2006. 958 с. 3. Zvit CERT-UA за 2014 рік, <http://cert.gov.ua/?p=2019>, 2015. 4. Peng Liu. Denial of Service Attacks – University Park, 2004. 5. Мартынюк И. Материалы технического тренинга «Построение безопасных сетей на оборудовании D-Link», <http://service.d-link.ua/sites/default/files/files/Security.zip>, Киев, 2012. 190с.

Поступила в редколлегию 21.02.2016

Рецензент: д-р техн. наук, проф. Безрук В.М.

Баранник Владимир Викторович, д-р техн. наук, профессор, начальник кафедры Харьковского университета Воздушных Сил. Научные интересы: кодирование и защита информации для передачи в телекоммуникационных системах. Адрес: Украина, 61000, Харьков, ул. Сумская, 77/79. E-mail: barannik_v_v@mail.ru.

Подлесный Сергей Анатоліевич, начальник отделения Харьковского университета Воздушных Сил. Научные интересы: кодирование и защита информации для передачи в телекоммуникационных системах. Адрес: Украина, 61000, Харьков, ул. Сумская, 77/79. E-mail: serg380638472732@gmail.com

Шульгин Сергей Сергеевич, соискатель Национального авиационного университета. Научные интересы: кодирования, семантической обработки изображений. Адрес: Украина, Киев, пр. Космонавта Комарова, 1.