Запорізький національний технічний університет

Радіоелектроніка Інформатика Управління

1(24)'2011

Науковий журнал

Виходить двічі на рік

Видається з березня 1999 року

Зареєстрований ***29 січня 2003 року***
Державним комітетом інформаційної політики,
телебачення та радіомовлення України.
Свідоцтво – серія ***КВ № 6904***

Засновник і видавник – *Запорізький національний технічний університет*

Запоріжжя, ЗНТУ
2011

ISSN 1607-3274

Постановою президії ВАК України № 1-05/4 від 26.05.2010 р. журнал «Радіоелектроніка, інформатика, управління» (скорочена назва – РІУ), який видається з 1999 року, включений до переліку наукових фахових видань України, в яких можуть публікуватися результати дисертаційних робіт на здобуття наукових ступенів доктора і кандидата технічних наук та фізико-математичних наук (фізика).

Журнал є донором журналу «Telecommunications and Radio Engineering», який видається в США.

Інтернет-сторінка журналу: http://journal.zntu.edu.ua/ric/index.php?page=index.

Статті, що публікуються в журналі, реферуються в базах даних та РЖ ВІНІТІ (Росія) і «Джерело» (Україна). Журнал РІУ міститься у міжнародній базі наукових видань Index Copernicus (http://journals.indexcopernicus.com/index.php), електронна копія журналу розміщена на сайті Національної бібліотеки України імені В. І. Вернадського НАН України у розділі «Наукова періодика України» за адресою: http://nbuv.gov.ua/portal/.

Журнал розповсюджується за Каталогом періодичних видань України (передплатний індекс – 22914).

РЕДАКЦІЙНА КОЛЕГІЯ

Рекомендовано до видання вченою радою Запорізького національного технічного університету, протокол № 4 від 22.11.2010 р.

Рукописи проходять незалежне рецензування з залученням провідних фахівців, за результатами якого редакційна колегія приймає рішення про опублікування.

Журнал зверстаний редакційно-видавничим відділом Запорізького національного технічного університету

Адреса редакції: 69063, м. Запоріжжя, вул. Жуковського, 64, ЗНТУ, редакція журналу «РІУ»

Тел: (061) 769-82-96 – редакційно-видавничий відділ
Факс: (061) 764-21-41
E-mail: rvv@zntu.edu.ua

З М І С Т

CONTENTS

РАДІОФІЗИКА

РАДИОФИЗИКА

RADIOPHYSICS

УДК 534.222.2

Никонова А. А.[1], Небеснюк О. Ю.[1], Шмалий С. Л.[1], Никонова З. А.[2]

[1]Канд. техн. наук, доцент Запорожской государственной инженерной академии
[2]Канд. техн. наук, профессор Запорожской государственной инженерной академии

ИССЛЕДОВАНИЕ МЕХАНИЗМОВ НЕСТАБИЛЬНОСТИ ХАРАКТЕРИСТИК МДП-СТРУКТУР

В статье приведены результаты исследования влияния термополевой обработки на характеристики МДП-структур.

Ключевые слова: заряд, термополевая обработка, МДП-структура.

ПОСТАНОВКА ПРОБЛЕМЫ

Известно, что зарядовое состояние МДП-структур определяется наличием в диэлектрике целого ряда зарядов, появляющихся как в нормальных условиях, так и в условиях повышенных температур и электрических полей [1]. Чтобы повысить стабильность их характеристик и установить вероятностные причины появления зарядов в диэлектрике, были исследованы механизмы нестабильности характеристик МДП-структур после термополевой обработки (ТПО).

Цель статьи – исследование влияния ТПО на характеристики МДП-структур.

АНАЛИЗ ПОСЛЕДНИХ ДОСТИЖЕНИЙ И ПУБЛИКАЦИЙ

Известно, что нестабильность эффективного заряда МДП-структур обусловлена генерацией или пространственным перераспределением заряда в диэлектрической пленке [2]. Это приводит к изменению величины зарядов в полупроводнике, а, следовательно, и к изменению поверхностного потенциала полупроводника. Исследования механизмов нестабильности МДП-систем показали, что чаще всего нестабильность связана с миграцией примесных ионов или переориентацией диполей в диэлектрике, накоплением носителей заряда на центрах захвата в объеме диэлектрика (рис. 1).

МАТЕРИАЛЫ И РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ

В реальных МДП-структурах существует много состояний и зарядов, которые влияют на идеальные кривые вольт-фарадных характеристик (ВФХ) этих структур.

Основная классификация этих зарядов и состояний следующая:

– фиксированные поверхностные заряды, которые локализуются вблизи поверхности полупроводника и не способны перемещаться под действием приложенного электрического поля;

– заряды подвижных ионов, способные перемещаться по объему диэлектрика под действием внешнего электрического поля;

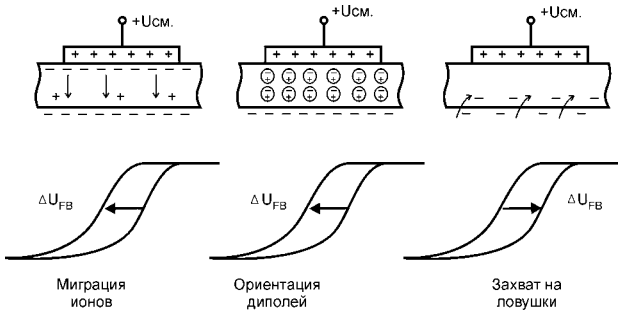


Рис. 1. Основные механизмы нестабильности МДП-структур

– поверхностные состояния (или состояния на границе раздела), которые определяются как энергетические уровни в запрещенной зоне на границе раздела диэлектрик-полупроводник, способные обмениваться зарядом с полупроводником в течение короткого времени.

Фиксированный заряд обладает следующими свойствами:

– плотность фиксированного заряда не зависит от величины поверхностного изгиба зон или от приложенного смещения;

– фиксированный заряд неподвижен, центры, ответственные за его образование, остаются неподвижными при проведении термополевой обработки;

– полярность фиксированного заряда всегда положительна, поэтому термически окисленная поверхность всегда *n*-типа, даже если нет загрязнения ионами щелочных металлов;

– фиксированные заряды всегда локализуются вблизи границы раздела диэлектрик-полупроводник.

Фиксированный заряд вызывает смещение вольт-фарадной характеристики на величину [3]:

$$\Delta U_{FB} = Q_{Ф} / C_{O,} \qquad (1)$$

где $Q_Ф$ – величина фиксированного заряда; C_O – емкость МДП-структуры.

Установлено, что при наличии этого заряда напряженность поля в диэлектрике МДП-структуры оказывается выше напряженности поля поверхности полупроводника. Для получения заданной величины поля приходилось прикладывать больший потенциал.

Одна из причин нестабильности МДП-структур – наличие в диэлектрике подвижных заряженных частиц, типа ионов щелочных металлов (Na, Li) или протонов. Обычно этот заряд вводится в диэлектрик при формировании последнего.

Эти частицы обладают относительно большой подвижностью, которая возрастает с увеличением температуры. Поэтому под действием электрического поля, особенно при повышенных температурах (100–300 °С), эти заряженные ионы могут легко перемещаться в диэлектрике, что приводит к изменению (нестабильности) во времени потенциала плоских зон.

Подвижность щелочных ионов в диэлектрике в значительной мере определяется величиной ионного радиуса элемента. При этом загрязнения Li более ощутимы, чем загрязнения Na и K. Однако вследствие того, что Na более распространен в природе, чем Li, именно Na определяет ионный заряд в диэлектрике.

При исследовании МДП-структур методом ВФХ присутствие подвижных примесей проявляется в сильном сдвиге ВФХ относительно теоретической (рис. 2), особенно после термополевой обработки (рис. 3).

В качестве объекта исследований были взяты две партии пластин с тестовыми (эталонными) МДП-структурами (полупроводник Si *p*-типа с кристаллографической ориентацией (111), диэлектрик – стекло толщиной 1000 Е).

I партия пластин. Спекание стеклопорошка проводилось при температуре 725 °С в среде кислорода с расходом 0,8–1,2 в течение 1,5–3 часов. Осаждение стекла проводилось электрофоретическим способом. Процесс осаждения длился 40 секунд, оплавление стекла происходило при температуре 950 °С ± 10 % в горячей зоне в течение 30–45 секунд в атмосфере кислорода (100–200 л/час).

II партия пластин. Помол стекла для этой партии проводился на мельнице Fritsch в стаканах из оплавленного электрокорунда. Режимы осаждения и оплавления стекла полностью соответствовали технологическому процессу нанесения стекла: время осаждения – 1, $I_{осажд.}$ – 80 А, $U_{осажд.}$ – 200 В, $T_{оплавл.}$ – 950 °С в атмосфере кислорода.

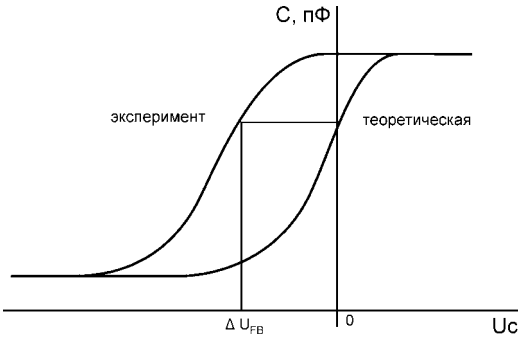


Рис. 2. Общий вид вольт-фарадной характеристики МДП-структур

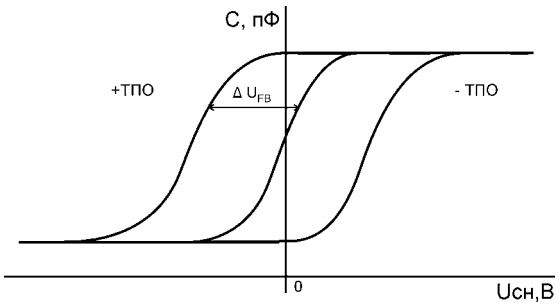


Рис. 3. Эффект действия термополевой обработки МДП-структур

После оплавления стекла пластины со структурами подвергались термополевой обработке, которая требует четкого контроля зарядов в диэлектрике. Для проведения исследований авторами была использована ранее разработанная мини-печь малой мощности с автоматизированным управлением параметрами полупроводниковых структур [4]. В ходе исследований установлено, что пластины с тестовыми структурами имели в исходном состоянии (при нулевом значении напряжения смещения) большие величины тангенса угла диэлектрических потерь. В связи с этим ВФХ на этих структурах не могли быть измерены.

Отдельные структуры при измерении ВФХ пробивались при подаче на них 50–100 В напряжения смещения. Образцы обеих партий (рис. 4, 5) имели отрицательный заряд. Эффективная плотность заряда, рассчитанная по ВФХ, составила $5–8 \cdot 10^{-8}$ см$^{-2}$.

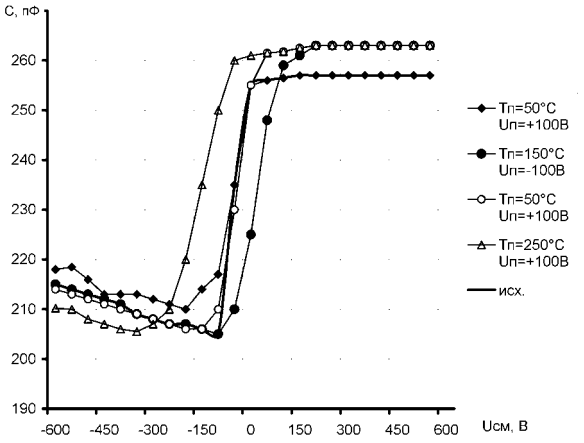


Рис. 4. Влияние термополевой обработки на зарядовое состояние тестовых МДП-структур. Партия I

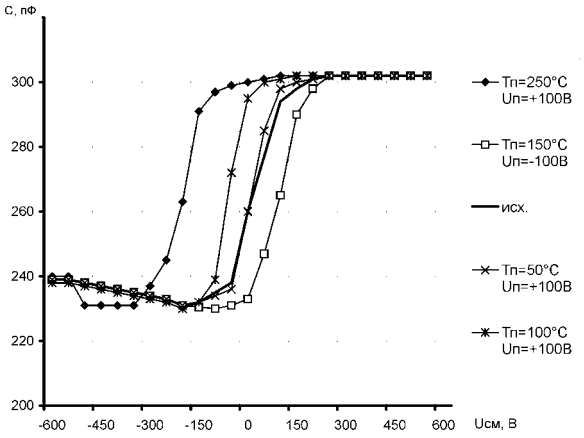


Рис. 5. Влияние термополевой обработки на зарядовое состояние тестовых МДП-структур. Партия II

Из экспериментальных данных (рис. 4, 5) следует, что на структурах, имеющих отрицательный заряд, после нагрева в обычной атмосфере до 250 °С наблюдается смещение потенциала плоских зон в область отрицательных напряжений, т. е. заряд таких структур становится положительным.

ВЫВОДЫ

В ходе исследований установлено:

– термополевая обработка оказывает существенное влияние на зарядовое состояние тестовых структур (инверсия заряда в диэлектрике), что приводит к нестабильности характеристик МДП-структур;

– при незначительной температуре и подаче напряжения на структуры наступает пробой и дальнейшие измерения на них становятся невозможными;

– распределение зарядов в диэлектрике МДП-структур оказывает значительное влияние на пороговое напряжение, что приводит к нестабильности их проходных характеристик. После ТПО при температуре до 250 °С наблюдалось существенное отклонение $U_{\text{пор.}}$ (±1–1,15 В) от граничных условий.

ПЕРЕЧЕНЬ ССЫЛОК

1. *Карабасов, Ю. С.* Новые материалы / Ю. С. Карабасов, В. Н. Анциферов, Ф. Ф. Бездудный, Л. Н. Белянчиков. – М. : МИСИС, 2002. – 736 с.
2. *Гуртов, В. А.* Электронные процессы в структурах металл – диэлектрик – полупроводник / В. А. Гуртов. – Петрозаводск, 1984. – 116 с.
3. *Nicollian, E. H.* MOS (Metal Oxide Semiconductor) Physics and Technology / E. H. Nicollian, J. R. Brews. – Wiley-Interscience, 1982. – 928 p.
4. *Ніконова, А. А.* Розробка міні-печі малої потужності з автоматизованим управлінням відпалу напівпровідникових структур / Ніконова З. А., Ситий М. Л., Шмалій С. Л. // Металургія : збірник наукових праць. – 2008. – Вип. 17. – С. 158–161.

Надійшла 06.09.2010
Після доробки 29.09.2010

Ніконова А. О., Небеснюк О. Ю., Шмалій С. Л., Ніконова З. А.

ДОСЛІДЖЕННЯ МЕХАНІЗМІВ НЕСТАБІЛЬНОСТІ ХАРАКТЕРИСТИК МДН-СТРУКТУР

У статті наведено результати дослідження впливу термополевої обробки на характеристики МДН-структур.

Ключові слова: заряд, термополева обробка, МДН-структура.

Nikonova A. A., Nebesnjuk O. J., Shmaly S. L., Nikonova Z. A.

INVESTIGATION OF MIS STRUCTURE CHARACTERISTICS INSTABILITY MECHANISMS

The results of investigation are presented in the paper showing the influence of thermal processing on characteristics of MIS structures.

Key words: charge, thermal processing, MIS structure.

УДК 537.874.6

Чумаченко Я. В.[1], Чумаченко В. П.[2]

[1]Канд. техн. наук, доцент Запорожского национального технического университета
[2]Д-р физ.-мат. наук, заведующий кафедрой Запорожского национального технического университета

О БЕСКОНЕЧНЫХ СИСТЕМАХ МЕТОДА ПРОИЗВЕДЕНИЯ ОБЛАСТЕЙ ДЛЯ ЗАДАЧ РАССЕЯНИЯ ВОЛН В ПЛОСКОСТНЫХ УЗЛАХ С СОЕДИНИТЕЛЬНОЙ ПОЛОСТЬЮ ПРЯМОУГОЛЬНОЙ ФОРМЫ

Изучаются свойства бесконечных систем линейных уравнений, возникающих при использовании метода произведения областей для определения характеристик рассеяния плоскостных волноводных трансформаторов с нагруженной соединительной полостью прямоугольной формы. На примере задачи о волноводном изломе показано, что в случае однородных условий Неймана на проводящих границах применение этого метода приводит к квазирегулярным системам.

Ключевые слова: бесконечные системы линейных уравнений, метод произведения областей, волноводные неоднородности.

ВВЕДЕНИЕ

Ряд широко используемых волноводных узлов имеет область связи, являющуюся общей частью пересекающихся под прямым углом бесконечных, полубесконечных или конечных волновых каналов. В качестве примеров укажем ответвители мощности, а также уголковые, Т-образные и крестообразные соединения волноводов (см. [1–5] и библиографию к ним). Подобные конфигурации возникают и в теории волноводов со сложным поперечным сечением [6]. Одним из методов, применяемых при анализе рассеяния волн в волноводных трансформаторах рассматриваемого типа, является метод произведения областей [7, 8], обеспечивающий адекватное представление поля внутри выпукло многоугольной соединительной полости. Согласно [7], искомая компонента поля в такой области задается суммой нескольких рядов по тригонометрическим функциям с разделенными переменными. Каждый из этих рядов тождественно удовлетворяет уравнению Гельмгольца внутри полости, а полнота используемых систем функций обеспечивает возможность выполнения требуемых граничных условий на ее контуре. Наложение условий сопряжения в апертурах области связи приводит к связанным бесконечным системам линейных алгебраических уравнений (СЛАУ) относительно неизвестных коэффициентов разложений поля в подобластях волноводного узла.

В [9, 10] было показано, что в случае однородных граничных условий Дирихле на стенках исследуемого объекта описанный подход порождает так называемые квазирегулярные СЛАУ. В настоящей работе на примере простейшего соединения описанного типа исследуются системы линейных уравнений, возникающие при нагруженной диэлектриком области связи и граничных условиях Неймана на контуре волноводного трансформатора. Полученные результаты применимы для соединений волноводов с магнитными стенками (ВМС), соединений плоскопараллельных волноводов (ППВ), а также Е-плоскостных соединений прямоугольных волноводов (ПВ). В последнем случае структура должна быть однородно заполненной, так как при наличии границ раздела сред задача рассеяния волн в таком узле уже не сводится к нахождению одной скалярной функции и требует отдельного рассмотрения [11]. Заметим также, что системы, близкие к исследуемым, появляются и при использовании метода частичных пересекающихся областей [12, 13].

ПОСТАНОВКА ЗАДАЧИ

Рассматриваемая конфигурация показана на рис. 1. Структура однородна вдоль оси z и имеет размер a в этом направлении ($a = \infty$ в случае ППВ). Внешний ее контур представляет собой сечение плоскостью $z = \text{const}$ идеальных электрических (ППВ, ПВ) или магнитных (ВМС) поверхностей. Области 1 и 2 соответствуют полубесконечным волноводам с поперечными размерами b и c соответственно. Прямоугольная область связи 3 или является незаполненной (ПВ), или нагружена диэлектриком (ППВ, ВМС) с относительной диэлектрической проницаемостью $\varepsilon = \varepsilon' - i\varepsilon''$. Узел возбуждается слева основной волной единичной амплитуды.

Задача состоит в отыскании ненулевой z-компоненты электромагнитного поля $H_z = u$ (ППВ), $H_z =$

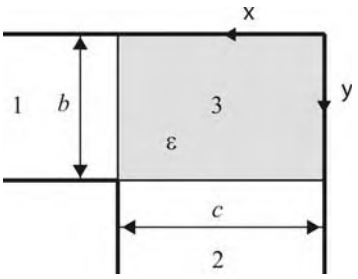


Рис. 1. Геометрия задачи

$= u\sin\frac{\pi z}{a}$ (ПВ) или $E_z = u$ (ВМС), где временная зависимость $e^{i\omega t}$ опущена. Функция u должна удовлетворять уравнению Гельмгольца, однородным граничным условиям Неймана на идеальных поверхностях, условиям непрерывности тангенциальных составляющих поля на границах частичных областей, условиям излучения на бесконечности и условиям ограниченности энергии поля, запасенной внутри любой конечной подобласти. Эти условия обеспечивают единственность решения исходной электродинамической задачи [14].

Обозначим значения u в областях 1, 2 и 3 через u_1, u_2 и u_3. Внутри своих областей функции u_j являются решениями уравнений

$$\Delta u_j + \chi_j^2 u = 0, \quad j = \overline{1,3} \tag{1}$$

со значениями параметров χ_j, определяемыми равенствами

$$\chi_1 = \chi_2 = \chi = k_0, \chi_3 = k_0\sqrt{\varepsilon} \text{ (ППВ, ВМС)} \tag{2}$$

или

$$\chi_1 = \chi_2 = \chi_3 = \chi = \sqrt{k_0^2 - \left(\frac{\pi}{a}\right)^2} \text{ (ПВ)}, \tag{3}$$

где $k_0 = \frac{2\pi}{\lambda}$, а λ – длина волны в свободном пространстве. Условия непрерывности тангенциальных компонент электромагнитного поля в апертурах области 3 приводят к соотношениям

$$u_1 = u_3, \frac{\partial u_1}{\partial x} = \kappa\frac{\partial u_3}{\partial x} \text{ при } x = c, \tag{4}$$

$$u_2 = u_3, \frac{\partial u_2}{\partial y} = \kappa\frac{\partial u_3}{\partial y} \text{ при } y = b. \tag{5}$$

Здесь

$$\kappa = \begin{cases} \chi^2/\chi_3^2 \text{ (ППВ)} \\ 1 \qquad \text{(ВМС, ПВ)} \end{cases}. \tag{6}$$

Тогда

$$u_1 = e^{\gamma_0^{(1)}(x-c)} + \sum_{n=0}^{\infty} A_n^{(1)}\cos\frac{n\pi y}{b}e^{-\gamma_n^{(1)}(x-c)}, \tag{7}$$

$$u_2 = \sum_{n=0}^{\infty} A_n^{(2)}\cos\frac{n\pi x}{c}e^{-\gamma_n^{(2)}(y-b)}, \tag{8}$$

где $A_n^{(1)}$ и $A_n^{(2)}$ – коэффициенты разложения, подлежащие определению, а

$$\gamma_n^{(1)} = \sqrt{\left(\frac{n\pi}{b}\right)^2 - \chi_1^2}, \ \gamma_n^{(2)} = \sqrt{\left(\frac{n\pi}{c}\right)^2 - \chi_2^2}. \tag{9}$$

Функцию u_3 будем искать в виде суммы двух рядов Фурье по косинусам

$$u_3 = \sum_{n=0}^{\infty} B_n^{(1)}\cos\frac{n\pi y}{b}\left[e^{\gamma_n^{(x)}(x-c)} + e^{-\gamma_n^{(x)}(x+c)}\right] + \\ + \sum_{n=0}^{\infty} B_n^{(2)}\cos\frac{n\pi x}{c}\left[e^{\gamma_n^{(y)}(y-b)} + e^{-\gamma_n^{(y)}(y+b)}\right], \tag{10}$$

$$\gamma_n^{(x)} = \sqrt{\left(\frac{n\pi}{b}\right)^2 - \chi_3^2}, \ \gamma_n^{(y)} = \sqrt{\left(\frac{n\pi}{c}\right)^2 - \chi_3^2}. \tag{11}$$

Представление (10) удовлетворяет однородным граничным условиям Неймана на идеальных стенках соединительной полости и имеет достаточную степень произвольности, чтобы обеспечить выполнение условий сопряжения (4), (5).

СИСТЕМА ЛИНЕЙНЫХ АЛГЕБРАИЧЕСКИХ УРАВНЕНИЙ

Подставив разложения (7) и (10) в условия (4) и воспользовавшись ортогональностью тригонометрических функций, мы получим

$$\delta_{0m} + A_m^{(1)} = (1 + e^{-2\gamma_m^{(x)}c})B_m^{(1)} + \sum_{n=0}^{\infty} d_{mn}^{(x)}B_n^{(2)}, \\ m = \overline{0,\infty}, \tag{12}$$

$$\delta_{0m}\gamma_0^{(1)} - \gamma_m^{(1)}A_m^{(1)} = \kappa\gamma_m^{(x)}(1 - e^{-2\gamma_m^{(x)}c})B_m^{(1)}, \\ m = \overline{0,\infty}, \tag{13}$$

где

$$d_{mn}^{(x)} = \frac{2(-1)^{m+n}\gamma_n^{(y)}(1 - e^{-2\gamma_n^{(y)}b})}{\varepsilon_m b\left[\gamma_n^{(y)2} + \left(\frac{m\pi}{b}\right)^2\right]}, \tag{14}$$

δ_{0m} – символ Кронекера, а $\varepsilon_m = \begin{cases} 2 \text{ при } m = 0 \\ 1 \text{ при } m \geq 1 \end{cases}$.

Граничные условия (5) приводят к уравнениям

$$A_m^{(2)} = (1 + e^{-2\gamma_m^{(y)}b})B_m^{(2)} + \sum_{n=0}^{\infty} d_{mn}^{(y)}B_n^{(1)}, \\ m = \overline{0,\infty}, \tag{15}$$

$$-\gamma_m^{(2)}A_m^{(2)} = \kappa\gamma_m^{(y)}(1 - e^{-2\gamma_m^{(y)}b})B_m^{(2)}, \\ m = \overline{0,\infty}, \tag{16}$$

$$d_{mn}^{(y)} = \frac{2(-1)^{m+n}\gamma_n^{(x)}(1-e^{-2\gamma_n^{(x)}c})}{\varepsilon_m c\left[\gamma_n^{(x)2}+\left(\frac{m\pi}{c}\right)^2\right]}. \qquad (17)$$

С помощью (13) и (16) $B_m^{(1)}$ и $B_m^{(2)}$ в (12) и (15) исключаются, что сводит рассматриваемую граничную задачу к парной СЛАУ относительно $A_m^{(1)}$ и $A_m^{(2)}$:

$$A_m^{(1)} + \sum_{n=0}^{\infty}\frac{\tilde{d}_{mn}^{(x)}}{\Delta_m^{(x)}}A_n^{(2)} = \frac{r\delta_{0m}}{\Delta_0^{(x)}},\ m = \overline{0,\infty}, \qquad (18)$$

$$A_m^{(2)} + \sum_{n=0}^{\infty}\frac{\tilde{d}_{mn}^{(y)}}{\Delta_m^{(y)}}A_n^{(1)} = \frac{\tilde{d}_{m0}^{(y)}}{\Delta_m^{(y)}},\ m = \overline{0,\infty}. \qquad (19)$$

Здесь

$$\tilde{d}_{mn}^{(x)} = \frac{2(-1)^{m+n}\gamma_m^{(x)}\gamma_n^{(2)}(1-e^{-2\gamma_m^{(x)}c})}{\varepsilon_m b\left[\left(\frac{n\pi}{c}\right)^2+\gamma_m^{(x)2}\right]}, \qquad (20)$$

$$\Delta_m^{(x)} = \gamma_m^{(1)}(1+e^{-2\gamma_m^{(x)}c}) + \kappa\gamma_m^{(x)}(1-e^{-2\gamma_m^{(x)}c}), \qquad (21)$$

$$r = \gamma_0^{(1)}(1+e^{-2\gamma_0^{(x)}c}) - \kappa\gamma_0^{(x)}(1-e^{-2\gamma_0^{(x)}c}), \qquad (22)$$

$$\tilde{d}_{mn}^{(y)} = \frac{2(-1)^{m+n}\gamma_m^{(y)}\gamma_n^{(1)}(1-e^{-2\gamma_m^{(y)}b})}{\varepsilon_m c\left[\left(\frac{n\pi}{b}\right)^2+\gamma_m^{(y)2}\right]}, \qquad (23)$$

$$\Delta_m^{(y)} = \gamma_m^{(2)}(1+e^{-2\gamma_m^{(y)}b}) + \kappa\gamma_m^{(y)}(1-e^{-2\gamma_m^{(y)}b}) \qquad (24)$$

и учтено, что

$$\gamma_n^{(y)2}+\left(\frac{m\pi}{b}\right)^2 = \left(\frac{n\pi}{c}\right)^2+\gamma_m^{(x)2}. \qquad (25)$$

СЛАУ (18), (19) является прямым следствием исходной краевой задачи и ее разрешимость следует из разрешимости краевой задачи. Условие конечности энергии поля в ограниченной области определяет класс числовых последовательностей, к которому должны принадлежать искомые амплитудные коэффициенты:

$$\sum_{n=0}^{\infty}|A_n^{(j)}|^2(n+1) < \infty,\ j = 1,2. \qquad (26)$$

Аналогичным условиям должны удовлетворять и коэффициенты $B_m^{(1)}$, $B_m^{(2)}$, как это следует из (13) и (16). Исходя из единственности решения краевой задачи и используя методику [15], можно доказать и единственность решения СЛАУ (18), (19) в пространстве последовательностей (26).

АНАЛИЗ МАТРИЧНЫХ КОЭФФИЦИЕНТОВ СЛАУ

Обозначим

$$a_{mn}^{(\xi)} = \frac{(m+1)\tilde{d}_{mn}^{(\xi)}}{(n+1)\Delta_m^{(\xi)}},\ \xi = x,y \qquad (27)$$

и введем новые неизвестные

$$\tilde{A}_m^{(j)} = A_m^{(j)}(m+1),\ j = 1,2. \qquad (28)$$

Тогда уравнения (18), (19) преобразуются к виду

$$\tilde{A}_m^{(1)} + \sum_{n=0}^{\infty}a_{mn}^{(x)}\tilde{A}_n^{(2)} = \frac{r\delta_{0m}}{\Delta_0^{(x)}},\ m = \overline{0,\infty}, \qquad (29)$$

$$\tilde{A}_m^{(2)} + \sum_{n=0}^{\infty}a_{mn}^{(y)}\tilde{A}_n^{(1)} = a_{m0}^{(y)},\ m = \overline{0,\infty}. \qquad (30)$$

Из (26) и (28) следует, что если последовательности $\tilde{A}_m^{(1)}$ и $\tilde{A}_m^{(2)}$ представляют решение краевой задачи, то они ограничены и сходятся к нулю.

Пусть далее

$$t_m = \left|\frac{\gamma_m^{(x)}(1-e^{-2\gamma_m^{(x)}c})}{\varepsilon_m\Delta_m^{(x)}}\right| \xrightarrow[m\to\infty]{} \frac{1}{|1+\kappa|} = t, \qquad (31)$$

$$\beta_{mn} = \frac{2(m+1)|\gamma_n^{(2)}|}{(n+1)b\left|\left(\frac{n\pi}{c}\right)^2+\gamma_m^{(x)2}\right|}, \qquad (32)$$

$$\beta_m = \sum_{n\le\alpha_c}\beta_{mn} \xrightarrow[m\to\infty]{} 0, \qquad (33)$$

где $\alpha_c = \frac{c\chi_2}{\pi}$. Заметив, что $\gamma_n^{(2)} < \frac{n\pi}{c}$ при $n > \alpha_c$, получим

$$\sum_{n=0}^{\infty}\beta_{mn} < \beta_m + \frac{2(m+1)c}{\pi b}\sum_{n=1}^{\infty}\frac{1}{\left|n^2+\left(\frac{c\gamma_m^{(x)}}{\pi}\right)^2\right|}. \qquad (34)$$

Предположим, что

$$m^2 > \left(\frac{b}{\pi}\right)^2\mathrm{Re}(\chi_3^2). \qquad (35)$$

Тогда

$$\left|n^2+\left(\frac{c\gamma_m^{(x)}}{\pi}\right)^2\right| \ge n^2+\left(\frac{c}{\pi}\right)^2\left[\left(\frac{m\pi}{b}\right)^2-\mathrm{Re}(\chi_3^2)\right] = n^2+s_m^2 \qquad (36)$$

и

$$\frac{2(m+1)c}{\pi b}\sum_{n=1}^{\infty}\frac{1}{\left|n^2+\left(\frac{c\gamma_m^{(x)}}{\pi}\right)^2\right|} \le \frac{2(m+1)c}{\pi b}\sum_{n=1}^{\infty}\frac{1}{n^2+s_m^2} \le \frac{(m+1)c}{bs_m} \xrightarrow[m\to\infty]{} 1, \qquad (37)$$

где принято во внимание, что

$$\sum_{n=1}^{\infty}\frac{1}{n^2+s^2} = \frac{\pi}{2s}\left(\mathrm{cths}\pi - \frac{1}{s\pi}\right)\ [16,\ \text{с. } 37] \qquad (38)$$

и

$$\mathrm{cth}p - \frac{1}{p} \le 1,\ 0 \le p < \infty. \qquad (39)$$

Учитывая (31)–(37), получим оценку

$$\sum_{n=0}^{\infty} |a_{mn}^{(x)}| = t_m \sum_{n=0}^{\infty} \beta_{mn} <$$

$$< t_m \left(\beta_m + \frac{(m+1)c}{bs_m} \right) \xrightarrow[m\to\infty]{} t. \qquad (40)$$

В случаях ПВ и ВМС $t = \frac{1}{2}$, а для ППВ $t < 1 - \delta$ ($\delta > 0$) для любого конечного значения диэлектрической проницаемости ε. Поэтому, исходя из определения предела, можно утверждать, что начиная с некоторого значения m,

$$\sum_{n=0}^{\infty} |a_{mn}^{(x)}| < 1 - \vartheta_1, \ \vartheta_1 > 0. \qquad (41)$$

Аналогично можно показать, что для достаточно больших m выполняется и условие

$$\sum_{n=0}^{\infty} |a_{mn}^{(y)}| < 1 - \vartheta_2, \ \vartheta_2 > 0. \qquad (42)$$

Заметим также, что

$$\sum_{n=0}^{\infty} |a_{mn}^{(\xi)}| < +\infty, \ \xi = x, y \qquad (43)$$

для всякого m, а правые части уравнений (29), (30) ограничены и стремятся к нулю.

Перенумеруем неизвестные

$$C_{2m-1} = A_{m-1}^{(1)}, \ C_{2m} = A_{m-1}^{(2)}, \ m = \overline{1, \infty} \qquad (44)$$

и перепишем в соответствующем порядке уравнения системы (29), (30). Учитывая установленные свойства матричных коэффициентов, можно утверждать, что полученная после этого бесконечная СЛАУ является квазирегулярной [17]. Пусть $M + 1$ – номер уравнения этой системы, начиная с которого оба условия (41) и (42) выполняются. Считая C_n, $n = \overline{1, M}$ заданными и отбросив первые M уравнений, получим вполне регулярную СЛАУ по отношению к неизвестным C_{M+n}, $n = \overline{1, \infty}$. Такая система имеет единственное ограниченное решение, которое может быть найдено методом редукции. Выразив из этой системы неизвестные со старшими номерами через первые неизвестные и подставив их в отбрасываемые равенства, мы сведем задачу к решению M уравнений относительно коэффициентов C_n, $n = \overline{1, M}$. Заметим, что если при построении приближения к решению вполне регулярной СЛАУ используется редуцированная система из N уравнений, то это эквивалентно усечению исходной СЛАУ до порядка $M + N$.

ЗАКЛЮЧЕНИЕ

На примере нагруженного излома волновода изучены свойства бесконечных СЛАУ, возникающих в результате применения метода произведения областей к исследованию рассеяния волн в волноводных трансформаторах с прямоугольной соединительной полостью и однородными условиями Неймана на проводящих границах. Установлено, что, как и в случае граничных условий Дирихле, этот метод приводит к квазирегулярным системам. Полученные результаты будут полезными при построении обоснованных алгоритмов для расчета технических характеристик волноводных узлов описанного типа.

СПИСОК ЛИТЕРАТУРЫ

1. *Шестопалов, В. П.* Резонансное рассеяние волн. Т. 2. Волноводные неоднородности / В. П. Шестопалов, А. А. Кириленко, Л. А. Рудь. – Киев : Наукова думка, 1986. – 216 с.
2. *Widarta, A.* Simple and accurate solutions of scattering coefficients of E-plane junctions in rectangular waveguides / A. Widarta, S. Kuwano and K. Kokubun // IEEE Transactions on Microwave Theory and Techniques. – 1995. – Vol. 43, Dec. – P. 2716–2718.
3. *Rebollar, J. M.* Fullwave analysis of three and four-port rectangular waveguide junctions / J. M. Rebollar, J. Esteban and J. E. Page // IEEE Transactions on Microwave Theory and Techniques. – 1994. – Vol. 42, Feb. – P. 256–263.
4. *Esteban, J.* Generalized scattering matrix of generalized two-port discontinuities: application to four-port and nonsymmetric six-port couplers / J. Esteban and J. M. Rebollar // IEEE Transactions on Microwave Theory and Techniques. – 1991. – Vol. 39, Oct. – P. 1725–1734.
5. *Liang, X.-P.* A rigorous three plane mode-matching technique for characterizing waveguide T-junctions, and its application in multiplexer design / X.-P. Liang, K. A. Zaki and A. E. Atia // IEEE Transactions on Microwave Theory and Techniques. – 1991. – Vol. 39, Dec. – P. 2138–2147.
6. *Заргано, Г. Ф.* Волноводы сложных сечений / Г. Ф. Заргано, В. П. Ляпин, В. С. Михалевский и др. – М. : Радио и связь, 1986. – 124 с.
7. *Chumachenko, V. P.* Efficient field representation for polygonal region / V. P. Chumachenko // Electronics Letters. – 2001. – Vol. 37, No. 19 – P. 1164–1165.
8. *Chumachenko, V. P.* Accurate analysis of waveguide junctions with rectangular coupling cavity / V. P. Chumachenko, E. Karaçuha and I. V. Petrusenko // Microwave and Optical Technology Letters. – 2001. – Vol. 31, Oct. – P. 305–308.
9. *Чумаченко, Я. В.* К обоснованию численного решения одной задачи рассеяния волн для нагруженного излома прямоугольного волновода / Я. В. Чумаченко, В. П. Чумаченко // Радіоелектроніка, інформатика, управління. – 2009. – № 2. – С. 32–34.
10. *Чумаченко, Я. В.* Исправления к статье «К обоснованию численного решения одной задачи рассеяния волн для нагруженного излома прямоугольного волновода» / Я. В. Чумаченко, В. П. Чумаченко // Радіоелектроніка, інформатика, управління. – 2011. – № 1. – С. 14.
11. *Kanellopoulos, V. N.* A complete E-plane analysis of waveguide junctions using the finite element method / V. N. Kanellopoulos and J. P. Webb // IEEE Transactions on Microwave Theory and Techniques. – 1990. – Vol. 38, Mar. – P. 290–295.
12. *Прохода, И. Г.* Расчет Н-плоскостного направленного ответвителя с учетом толщины общей стенки между волноводами / И. Г. Прохода, В. И. Лозяной, В. П. Чумаченко // Изв. вузов. Радиофизика. – 1974. – Т. 17, № 8. – С. 1214–1218.

13. *Yakovlev, A. B.* Analysis of microstrip discontinuities using method of integral equations for overlapping regions / A. B. Yakovlev and A. B. Gnilenko // IEE Proceedings. Microwaves, Antennas and Propagation. – 1997. – Vol. 144, Dec. – P. 449–457.
14. *Хенл, Х.* Теория дифракции / Х. Хенл, А. Мауэ, К. Вестпфаль. – М. : Мир, 1964. – 428 с.
15. *Шестопалов, В. П.* Матричные уравнения типа свертки в теории дифракции / В. П. Шестопалов, А. А. Кириленко, С. А. Масалов.– Киев : Наукова думка, 1984. – 296 с.
16. *Градштейн, И. С.* Таблицы интегралов, сумм, рядов и произведений / И. С. Градштейн, И. М. Рыжик. – М. : Наука, 1971. – 1108 с.
17. *Канторович, Л. В.* Приближенные методы высшего анализа / Л. В. Канторович, В. И. Крылов. – М.-Л. : Физматгиз, 1962. – 708 с.

Надійшла 04.11.2010

Чумаченко Я. В., Чумаченко В. П.

ПРО НЕСКІНЧЕННІ СИСТЕМИ МЕТОДУ ДОБУТКУ ОБЛАСТЕЙ ДЛЯ ЗАДАЧ РОЗСІЮВАННЯ ХВИЛЬ В ПЛОЩИННИХ ВУЗЛАХ ЗІ З'ЄДНУВАЛЬНОЮ ПОРОЖНИНОЮ ПРЯМОКУТНОЇ ФОРМИ

Вивчаються властивості нескінченних систем лінійних рівнянь, які виникають при використанні методу добутку областей для знаходження характеристик розсіювання площинних хвилевідних трансформаторів з навантаженою з'єднувальною порожниною прямокутної форми. На прикладі задачі про злам хвилеводу показано, що у випадку однорідних умов Неймана на провідних межових поверхнях застосування цього методу приводить до квазірегулярних систем.

Ключові слова: нескінченні системи лінійних рівнянь, метод добутку областей, хвилевідні неоднорідності.

Chumachenko Ya. V., Chumachenko V. P.

ON LINEAR INFINITE SYSTEMS OF DOMAIN-PRODUCT TECHNIQUE FOR WAVE SCATTERING PROBLEMS IN PLANAR WAVEGUIDE JUNCTIONS WITH RECTANGULAR CONNECTING CAVITY

Properties of infinite systems of linear equations that occur when applying the domain-product technique to scattering problems for planar waveguide transformers with a loaded rectangular connecting cavity are studied. By the example of a waveguide bend, it is shown that in case of homogeneous Neumann conditions at conducting boundaries the method results in quasiregular systems.

Key words: linear infinite systems, domain-product technique, waveguide discontinuities.

УДК 537.874.6

Чумаченко Я. В.[1], Чумаченко В. П.[2]

[1]Канд. техн. наук, доцент Запорожского национального технического университета

[2]Д-р физ.-мат. наук, заведующий кафедрой Запорожского национального технического университета

ИСПРАВЛЕНИЯ К СТАТЬЕ «К ОБОСНОВАНИЮ ЧИСЛЕННОГО РЕШЕНИЯ ОДНОЙ ЗАДАЧИ РАССЕЯНИЯ ВОЛН ДЛЯ НАГРУЖЕННОГО ИЗЛОМА ПРЯМОУГОЛЬНОГО ВОЛНОВОДА»

В работе [1] авторы обнаружили ошибку, которая, однако, не повлияла на полученные результаты.

Если k_0 таково, что для $n \le N$ $\gamma_n^{(y)} = i\beta_n$ (β_n – действительная величина), то $\left|e^{-2\gamma_n^{(y)}d} - 1\right| \le 2$ при $n \le N$ и $\left|e^{-2\gamma_n^{(y)}d} - 1\right| \le 1$ при $n > N$. Введем обозначения

$$\alpha = \frac{m^2 c}{d^2 \gamma_m^{(x)}} \sum_{n=1}^{N} \frac{1}{n^2 + \left(\gamma_m^{(x)} \frac{c}{\pi}\right)^2} \xrightarrow[m\to\infty]{} 0,$$

$$A = \frac{m^2 c}{d^2 \gamma_m^{(x)}} \sum_{n=1}^{\infty} \frac{1}{n^2 + \left(\gamma_m^{(x)} \frac{c}{\pi}\right)^2}.$$

Тогда в [1] оценка (14) с учетом соотношений (15)–(17) должна иметь вид

$$\sum_{n=1}^{\infty} \left| \frac{\widetilde{d}_{mn}^{(x)}}{\Delta_m^{(x)}} \right| < \sum_{n=1}^{\infty} \frac{\left|\widetilde{d}_{mn}^{(x)}\right|}{2\gamma_m^{(x)}} \le \alpha + A \le$$

$$\le \alpha + \frac{1}{2} \frac{1}{1 - \left(\frac{\chi d}{m\pi}\right)^2} \xrightarrow[m\to\infty]{} \frac{1}{2}.$$

Это означает, что, начиная с некоторого значения m, выполняется неравенство (19). Справедливость неравенства (21) для достаточно больших m устанавливается аналогично. Отсюда следует, что итоговая СЛАУ является квазирегулярной и известным образом может быть сведена к конечной системе после применения метода усечения к уравнениям, для которых выполняются оба условия (19) и (21).

СПИСОК ЛИТЕРАТУРЫ

1. *Чумаченко, Я. В.* К обоснованию численного решения одной задачи рассеяния волн для нагруженного излома прямоугольного волновода / Я. В. Чумаченко, В. П. Чумаченко // Радіоелектроніка, інформатика, управління. – 2009. – № 2. – С. 32–34.

РАДІОЕЛЕКТРОНІКА ТА ТЕЛЕКОМУНІКАЦІЇ

РАДИОЭЛЕКТРОНИКА И ТЕЛЕКОММУНИКАЦИИ

RADIO ELECTRONICS AND TELECOMMUNICATIONS

УДК 621.396.6

Єфіменко А. А.

Канд. техн. наук, доцент Одеського національного політехнічного університету

ВИБІР ОПТИМАЛЬНИХ КОНСТРУКЦІЙ МІЖБЛОЧНИХ ЕЛЕКТРИЧНИХ З'ЄДНАНЬ ДЛЯ ЕЛЕКТРОННИХ ЗАСОБІВ

Запропоновано моделі та алгоритми вибору конструкцій електричних з'єднань за критеріями вартості і трудомісткості, показниками якості та ефективності, які дозволяють оптимізувати процес розробки електронних засобів.

Ключові слова: електричні з'єднання, міжконтактні з'єднання, оптимізація вартості, електроні пристрої.

ВСТУП

Існує досить велика кількість варіантів конструктивно-технологічного виконання міжблочних електричних з'єднань електронних засобів (ЕЗ), які розглядаються у двох нерозривно зв'язаних між собою аспектах – міжконтактні з'єднання і контактні з'єднання (власне контактування). Різноманіття варіантів об'єднання міжконтактних з'єднань (видів електричних з'єднань) і контактних з'єднань (методів електричних з'єднань), які мають різні техніко-економічні показники, приводить до необхідності визначення оптимальних варіантів конструкцій електричних з'єднань на базі впровадження математичного апарату та широкого застосування засобів обчислювальної техніки і є актуальною задачею.

У свою чергу, обумовлюється актуальність розвитку методологічних аспектів проектування, що включають у себе виявлення і дослідження комплексу практично необхідних вимог до перспективних електричних з'єднань, розробку системи класифікації їхніх параметрів і показників якості, математичну постановку задач синтезу оптимальних варіантів електричних з'єднань. Важливим є також ефективність методів проектування, що впроваджуються, – вони повинні забезпечувати високу достовірність результатів при незначних матеріальних та часових затратах.

Дослідження в напрямку вирішення даних проблем не носять систематичного характеру і останнім часом обмежуються розробкою нових видів та методів електричних з'єднань без створення об'єктивних засобів їх аналізу та оптимального використання, наприклад [1–5]. Попередні дослідження (деякі з них представлені в [6–10]) потребують врахування розвитку теорії та практики створення ЕЗ і використання сучасних можливостей обчислювальної техніки та програмних засобів проектування. Крім того, слід враховувати, що розробка оптимальної конструкції електричних з'єднань являє собою досить складну науково-технічну задачу, тому що вибір необхідно вести за багатьма критеріями та при різних обмеженнях, що накладаються на умови задачі та елементи її рішення. Це створює передумови для розробки та використання достатньо великої множини моделей оптимізації [11].

Метою даної статті є розробка моделей та алгоритмів вибору (синтезу) і оптимізації конструкцій електричних з'єднань в електронних засобах, які можна було б використовувати на ранніх стадіях проектування. При цьому доцільно використовувати різні критерії і, відповідно, моделі з метою надання розробникам ЕЗ різних можливостей щодо деталізації початкових даних і достовірності результатів оптимізації.

З метою врахування всіх можливих варіантів конструкцій електричних з'єднань дуже важливим є розгляд всіх їх складових елементів, а також тих інфраструктур, де вони будуть використовуватись.

Зважаючи на суттєвий взаємовплив електричних з'єднань та інших частин ЕЗ, можна зробити висновок, що електричні з'єднання є важливою частиною електронних засобів і потребують найбільш пильної уваги при їх проектуванні за умов отримання оптимальних результатів. Вирішуючи задачу розробки електричних з'єднань у складі ЕЗ, можна керуватися схемою, представленою на рис. 1, на якій конкретизовано деякі роботи (штрих-пунктирною лінією окреслені блоки, що стосуються розробки електричних з'єднань). На схемі використані позначення: ТЗ – технічне завдання; ЭЗ – схема електрична принципова; ПЭЗ – перелік елементів до схеми електричної принципової; НК – несучі конструкції; КД – конструкторська документація.

Відповідно до розглянутої схеми важливою частиною створення електричних з'єднань є вибір їх структури (варіантів конструкції) та її оптимізація.

Вибір оптимального варіанту конструкції електричних з'єднань доцільно вести за такими напрямками:

1) мінімальна вартість;

2) мінімальна трудомісткість;

3) максимальні показники якості;

4) максимальні показники ефективності.

Ці варіанти не рівноцінні як з точки зору затрат на їх реалізацію, так і з позицій достовірності результатів оптимізації. Звичайно і використання тих чи інших варіантів є прерогативою розробників ЕЗ і залежать від умов, в яких вони діють.

Нижче наведено математичні моделі оптимізації, що відповідають розглянутим напрямкам.

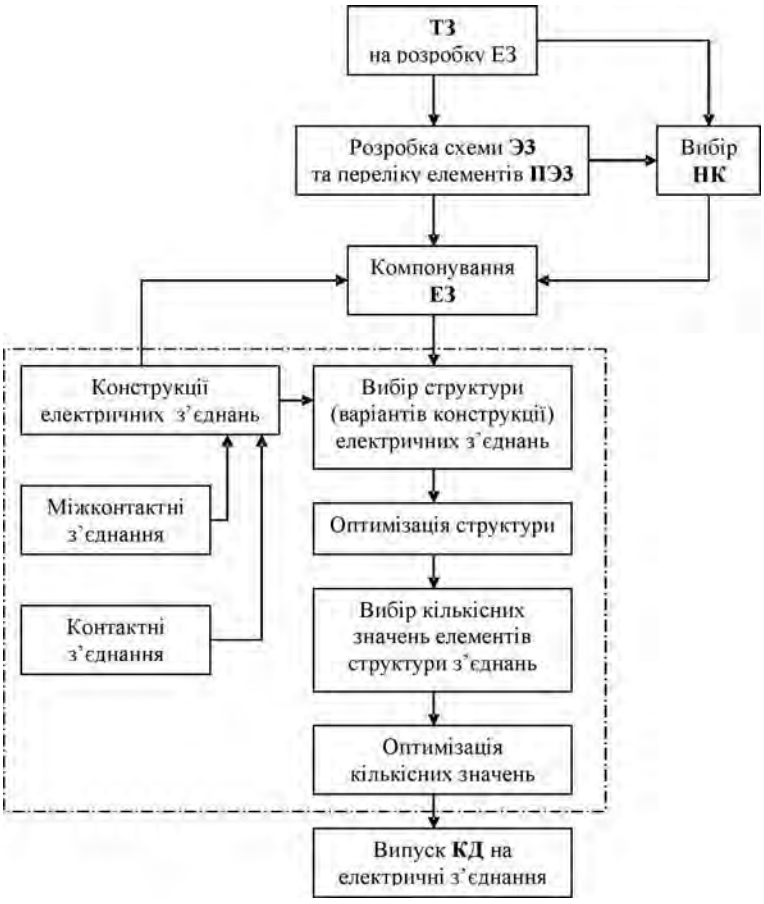


Рис. 1. Процес розробки ЕЗ з конкретизацією деяких робіт зі створення електричних з'єднань

МОДЕЛІ ВИБОРУ КОНСТРУКЦІЙ ЕЛЕКТРИЧНИХ З'ЄДНАНЬ

1. Мінімізація вартості електричних з'єднань

В цьому випадку модель вибору структури (варіанта конструкції) електричних з'єднань ЕЗ може бути представлена як

$$K_1 = \min \sum_{i=1}^{n} C_i \qquad (1)$$

при обмеженнях $a \subset a_{\text{доп}}$, $r \subset r_{\text{доп}}$, $h \subset h_{\text{доп}}$, де a, r, h – відповідно типи міжконтактних, контактних з'єднань і елементів кріплення, які лежать в області припустимих типів $a_{\text{доп}}$, $r_{\text{доп}}$, $h_{\text{доп}}$; n – кількість складових частин конструкції електричних з'єднань; $C = \{M, Q, P\}$ – вартість електричних з'єднань; $M \subset \subset \{M_1, M_2, \ldots, M_a, \ldots, M_s\}$ – множина вартостей міжконтактних з'єднань M, що складається з вартостей окремих типів міжконтактних з'єднань; s – кількість типів міжконтактних з'єднань; $Q \subset \{Q_1, Q_2, \ldots, Q_r, \ldots, Q_t\}$ – множина вартостей контактних з'єднань Q, що складається з вартостей окремих типів контактних з'єднань; t – кількість типів контактних з'єднань; $P \subset \{P_1, P_2, \ldots, P_h, \ldots, P_l\}$ – множина вартостей елементів кріплення P, що складається з вартостей окремих типів кріплення; l – кількість типів елементів кріплення.

У свою чергу

$$M_a = \sum_{k=1}^{d} m_k,\ Q_r = \sum_{v=1}^{z} q_v,\ P_h = \sum_{w=1}^{q} p_w,$$

де m_k – вартість окремих складових міжконтактних з'єднань; q_v – вартість окремих складових контактних з'єднань; p_w – вартість окремих складових елементів кріплення.

2. Мінімізація трудомісткості виготовлення електричних з'єднань

Модель вибору має вигляд

$$K_2 = \min \sum_{i=1}^{n} T_i \qquad (2)$$

при обмеженнях $a \subset a_{\text{доп}}$, $r \subset r_{\text{доп}}$, $h \subset h_{\text{доп}}$, де T_i – трудомісткість i-го типу електричних з'єднань.

Ця модель є спрощеною модифікацією попередньої моделі у зв'язку з тим, що трудомісткість є частиною всіх затрат, тобто повної вартості. В деяких випадках зручніше користуватись саме цим показником.

3. Максимізація показників якості

Модель вибору оптимальної структури (варіанту конструкції) електричних з'єднань ЕЗ за показниками якості має вигляд

$$K_3 = \max \sum_{j=1}^{b} K_j \qquad (3)$$

при обмеженнях $a \subset a_{\text{доп}}$, $r \subset r_{\text{доп}}$, $h \subset h_{\text{доп}}$, де b – кількість часткових показників якості K_j електричних з'єднань.

Як часткові можуть бути використані такі показники якості [12]:

– коефіцієнт об'єму конструкції електричних з'єднань

$$K_V = 1 - \frac{\sum_{i=1}^{n} V_i l_i + \sum_{i=1}^{n} V_{ki} n_{ki}}{V}, \qquad (4)$$

де V_i – об'єм одиниці довжини i-го з'єднання; V_{ki} – об'єм контакту i-го з'єднання; V– об'єм всього виробу; l_i – довжина i-го з'єднання; n_{ki} – число контактів i-го з'єднання; n – число з'єднань у виробі;

– показник питомого об'єму конструкції електричних з'єднань

$$V_{\text{пит}} = \frac{\sum_{i=1}^{n} V_i l_i + \sum_{i=1}^{n} V_{ki} n_{ki}}{n}; \qquad (5)$$

– коефіцієнт маси конструкції електричних з'єднань

$$K_m = 1 - \frac{\sum_{i=1}^{n} m_i l_i + \sum_{i=1}^{n} m_{ki} n_{ki}}{M}, \qquad (6)$$

де m_i – маса одиниці довжини i-го з'єднання; m_{ki} – маса контакту i-го з'єднання; M – маса всього виробу;

– показник питомої маси конструкції електричних з'єднань

$$m_{\text{пит}} = \frac{\sum_{i=1}^{n} m_i l_i + \sum_{i=1}^{n} m_{ki} n_{ki}}{n}; \qquad (7)$$

– показник питомого напрацювання на відмову конструкції електричних з'єднань

$$t_{\text{пит}} = \frac{1}{\left(\sum_{i=1}^{n} \lambda_i + \sum_{j=1}^{d} \lambda_j \right) \cdot n}, \qquad (8)$$

де λ_i – інтенсивність відмови i-го з'єднання; λ_j – інтенсивність відмови j-го контакта; d – кількість контактів у виробі;

– показник приведеного часу затримки передачі (обробки) інформації при використанні заданої конструкції електричних з'єднань

$$\tau_{\text{пр}} = \sum_{i=1}^{n} \tau_i / \sum_{i=1}^{n} l_i, \qquad (9)$$

де τ_i – затримка сигналу в i-му з'єднанні тракту передачі (обробки) інформації;

– коефіцієнт автоматизації виконання електромонтажу

$$K_a = n_a / n, \qquad (10)$$

де n_a – кількість з'єднань, що виконуються автоматизованими методами.

Для рішення задачі вибору оптимального варіанту конструкції електричних з'єднань використаємо теорію пасивних ігор [13]. Ця теорія добре підходить для вирішення такого класу задач з неповною початковою інформацією.

Складемо ігрову матрицю (табл. 1). Рядки матриці відповідають варіантам конструкції електричних з'єднань Y_i, а стовпці – їх частковим показникам якості K_j. В комірки ігрової матриці заносяться розраховані значення часткових показників якості Π_{ij} за різними варіантами конструкції Y_i.

Серед часткових показників якості є такі, що максимізуються, і такі, що мінімізуються. Для наведених вище показників максимізуються K_v, K_m, $t_{\text{пит}}$, K_a і мінімізуються $V_{\text{пит}}$, $m_{\text{пит}}$, $\tau_{\text{пр}}$. Для об'єднання (згортки) всіх часткових показників в інтегральний, наприклад, у вигляді суми, потрібно табл. 1 фактично розділити на дві: одна – з матрицею значень показників якості, що максимізуються $[\Pi \max_{ij}]$, друга – з матрицею значень показників якості, що мінімізуються $[\Pi \min_{ij}]$. В свою чергу, для обох таблиць кількість варіантів конструкції однакова, тобто $\boldsymbol{i}$ лежить у межах від 1 до $\boldsymbol{u}$, а кількість часткових показників якості в загальному випадку може бути різною: для максимізуємих $\boldsymbol{j}$ лежить у межах від 1 до $\boldsymbol{c}$, для мінімізуємих – від 1 до $\boldsymbol{e}$, при цьому $\boldsymbol{c} + \boldsymbol{e} = \boldsymbol{b}$.

Таблиця 1

Варіант конструкції	K_1	K_2	…	K_j	…	K_b
Y_1	Π_{11}	Π_{12}.	…	Π_{1j}	…	Π_{1b}.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
Y_i	Π_{i1}	Π_{i2}	…	Π_{ij}	…	Π_{ib}
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
Y_u	Π_{u1}	Π_{u2}	…	Π_{uj}	…	Π_{ub}

Часткові показники якості, що використовуються, є різними за фізичною природою і можуть бути розмірними та такими, що не мають розмірності. При цьому значення кожного $\boldsymbol{j}$-го показника можуть суттєво відрізнятися між собою за величиною. У зв'язку з цим потрібно виконати нормування значень показників якості, використовуючи таке співвідношення:

$$\Pi_{ij\text{н}} = \Pi_{ij} / \max \Pi_j, \qquad (11)$$

де Π_{ij} – значення показників якості в кожній із двох матриць; $\max \Pi_j$ – максимальне значення $\boldsymbol{j}$-го показника якості, тобто максимальне значення в кожному стовпці.

У результаті отримуємо дві пронормовані матриці $[\Pi \max_{ij\text{н}}]$ та $[\Pi \min_{ij\text{н}}]$.

Часткові показники якості, як такі, що максимізуються, так і такі, що мінімізуються, по-різному впливають на інтегральний показник якості, за яким визначається оптимальний варіант конструкції електричних з'єднань. Ступінь впливу можливо врахувати за допомогою коефіцієнтів вагомості, які можна визначити, наприклад, експертним шляхом. З цією метою елементи обох матриць (значення часткових показників) помножуються на коефіцієнти вагомості. В результаті отримуємо пронормовані зважені матриці $[\Pi \max^*_{ij\text{н}}]$ та $[\Pi \min^*_{ij\text{н}}]$.

Наступним кроком є згортка значень показників якості для кожного варіанта конструкції електричних з'єднань в межах кожної з двох матриць – за максимізуємими та мінімізуємими показниками:

$$K_{3\max i} = \sum_{j=1}^{c} \Pi \max^*_{ij\text{н}}; \; K_{3\min i} = \sum_{j=1}^{e} \Pi \min^*_{ij\text{н}}. \qquad (12)$$

Для визначення оптимального варіанту конструкції електричних з'єднань виконуємо згортку загальних значень максимізуємих та мінімізуємих показників якості для кожного варіанта у вигляді частки

$$K_{3i} = \sum_{j=1}^{c} \Pi \max^*_{ij\text{н}} / \sum_{j=1}^{e} \Pi \min^*_{ij\text{н}}. \qquad (13)$$

Оптимальним буде той варіант, який дає максимальне відношення $K_{3i} \to \max$.

4. Максимізація показників ефективності

У ряді випадків доцільно використовувати не просто показники якості, а показники ефективності – відношення показників якості до вартості варіанту конструкції електричного з'єднання. Це дає змогу оцінювати не тільки технічні характеристики,

а й економічні, тобто враховувати загалом співвідношення «якість/вартість».

Модель вибору оптимального варіанту конструкції електричних з'єднань ЕЗ за показниками ефективності має вигляд

$$K_4 = \max \sum_{j=1}^{b} E_j \qquad (14)$$

при тих же обмеженнях $a \subset a_{\text{доп}}$, $r \subset r_{\text{доп}}$, $h \subset h_{\text{доп}}$, де Ej – часткові показники ефективності.

Для рішення задачі оптимізації представляємо всі часткові показники якості у вигляді показників ефективності (виграшів), які завжди максимізуються, за допомогою співвідношень:

– для максимізуємих часткових показників

$$E_j = \frac{K_j}{C_i},$$

де C_i– вартість ***i***-го варіанта конструкції електричних з'єднань;

– для мінімізуємих часткових показників

$$E_j = \frac{1}{K_j \cdot C_i}.$$

Для розрахунку елементів матриці ефективності (виграшів) слід використовувати такі співвідношення, що витікають із попередніх:

$$E_{ij} = \frac{\Pi \max_{ij}}{C_i} \text{ та } E_{ij} = \frac{1}{\Pi \min_{ij} \cdot C_i}. \qquad (15)$$

Розрахувавши таким чином показники ефективності, побудуємо матрицю виграшів (табл. 2), в якій всі показники максимізуються.

Таблиця 2

Варіант конструкції	K_1	K_2	…	K_j	…	K_b
Y_1	E_{11}	E_{12}.	…	E_{1j}	…	E_{1b}.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
Y_i	E_{i1}	E_{i2}	…	E_{ij}	…	E_{ib}
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
Y_u	E_{u1}	E_{u2}	…	E_{uj}	…	E_{ub}

Як і у випадку використання попередньої моделі, для визначення оптимального варіанту конструкції електричних з'єднань слід виконати такі операції:

– нормування значень показників ефективності за таким співвідношенням:

$$E_{ij\text{н}} = E_{ij} / \max E_j, \qquad (16)$$

де $E_{ij\text{н}}$ – нормоване значення показника ефективності ($E_{ij\text{н}} \le 1$); $\max E_j$ – максимальне значення показника ефективності по кожному частковому показнику якості (в кожному стовпці матриці);

– уведення коефіцієнтів вагомості та отримання пронормованої зваженої матриці $[E^*_{ij\text{н}}]$;

– згортка часткових показників ефективності в інтегральний у вигляді суми для кожного з варіантів конструкції електричних з'єднань:

$$K_{4i} = \sum_{j=1}^{b} E^*_{ij\text{н}}; \qquad (17)$$

– визначення оптимального варіанта конструкції як

$$K_{4i} \to \max.$$

АЛГОРИТМИ ВИБОРУ ОПТИМАЛЬНИХ КОНСТРУКЦІЙ ЕЛЕКТРИЧНИХ З'ЄДНАНЬ

У відповідності до представлених моделей розроблено алгоритми (рис. 2, 3), які дозволяють в різних умовах проводити вибір конструкцій електричних з'єднань.

Враховуючи наявність однакових початкових даних, баз даних та операцій, що виконуються, алгоритми побудовані таким чином, що чотири моделі вибору реалізовані двома алгоритмами.

1. Алгоритм вибору за критеріями трудомісткості та вартості

Алгоритм представлений на рис. 2, при цьому визначається мінімальна трудомісткість чи вартість конструкції електричних з'єднань. Вибір конструкції відбувається на основі порівняння трудомісткості виготовлення їх варіантів, якщо цього достатньо (алгоритм використовується частково), або порівняння повної вартості варіантів (алгоритм використовується повністю).

Блок 1. Процес вибору варіанта електромонтажу передбачає наявність цих варіантів. Тому насамперед потрібно в ескізному вигляді розробити ці варіанти. Для цього необхідні початкові дані (**блок 2**) – як мінімум, схема електрична принципова (ЭЗ), несуча конструкція (НК), яка буде використана для виробу, що розробляється, та компонувальний ескіз. Звичайно, на момент розробки виробу відомі набори можливих варіантів конструкції електричних з'єднань, а також частина їх, що рекомендуються для переважного використання з урахуванням вимог до виробу, умов його експлуатації, а також рівня розвитку технологій електромонтажу (**блок 3**). Аналіз цих даних дозволяє визначити обмеження до математичної

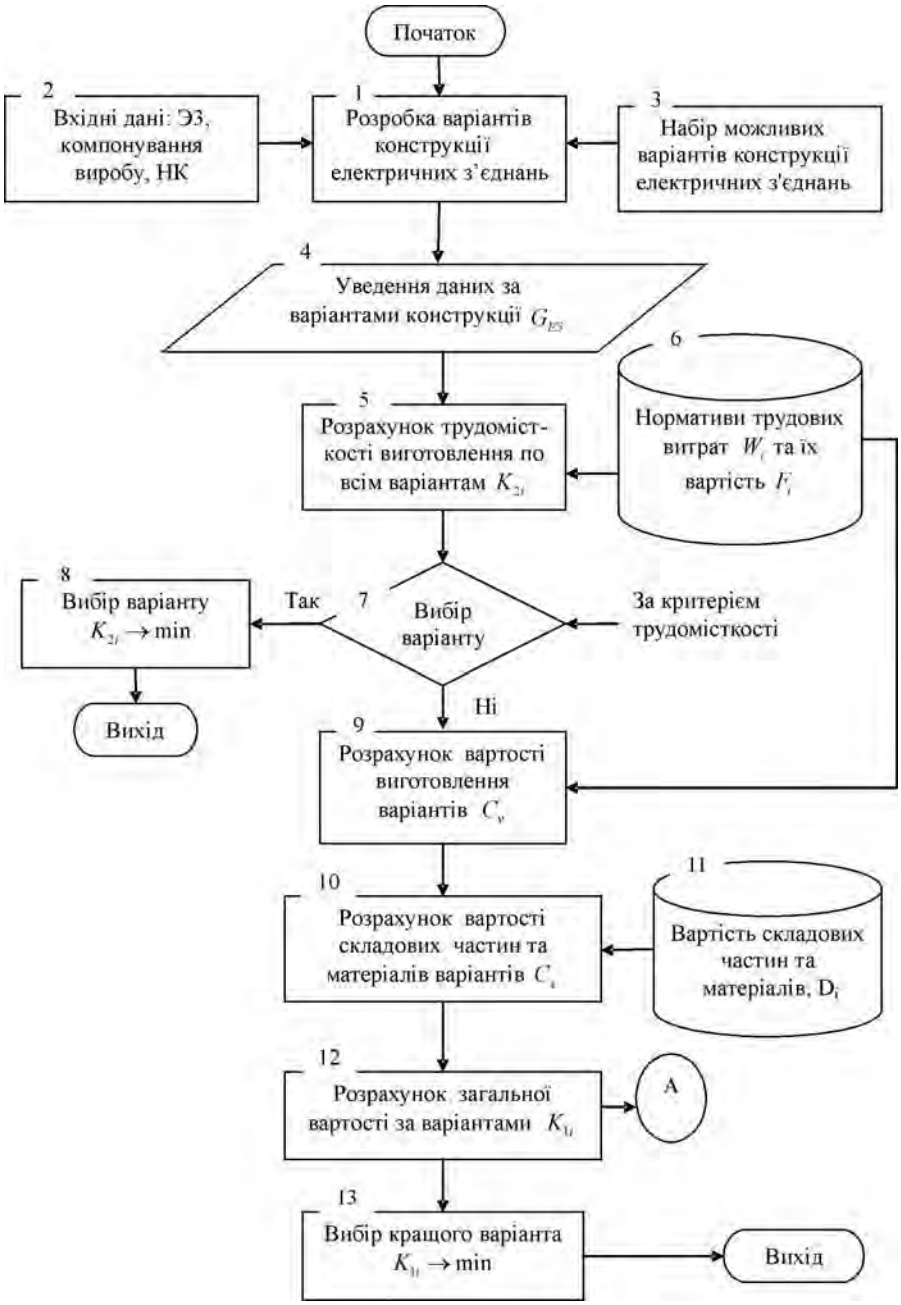


Рис. 2. Алгоритм вибору варіанта конструкції електричних з'єднань за критеріями трудомісткості та вартості

моделі вибору оптимального варіанту конструкції електричних з'єднань ЕЗ, що розробляється.

Блок 4. Для кожного із розроблених варіантів електромонтажу уводяться кількісні дані:

– типи проводів та кабелів;

– довжина проводів та кабелів за типом;

– інші матеріали;

– кількість роз'ємів;

– кількість підготовчих операцій (нарізка, лудіння, зачищення та ін.);

– кількість операцій монтажу;

– типи елементів кріплення та їх кількість;

– інші дані.

Блок 5. Для всіх прийнятих до розгляду варіантів конструкції електричних з'єднань проводиться розрахунок трудомісткості виконання електромонтажу з урахуванням всіх технологічних операцій, у тому числі підготовчих. Для цього крім даних, що надходять з блоку 4, використовуються нормативи трудових витрат, що знаходяться в базі даних (**блок 6**).

Блок 7. Визначається напрямок подальших розрахунків. Якщо зазначено вибір електромонтажу за кри-

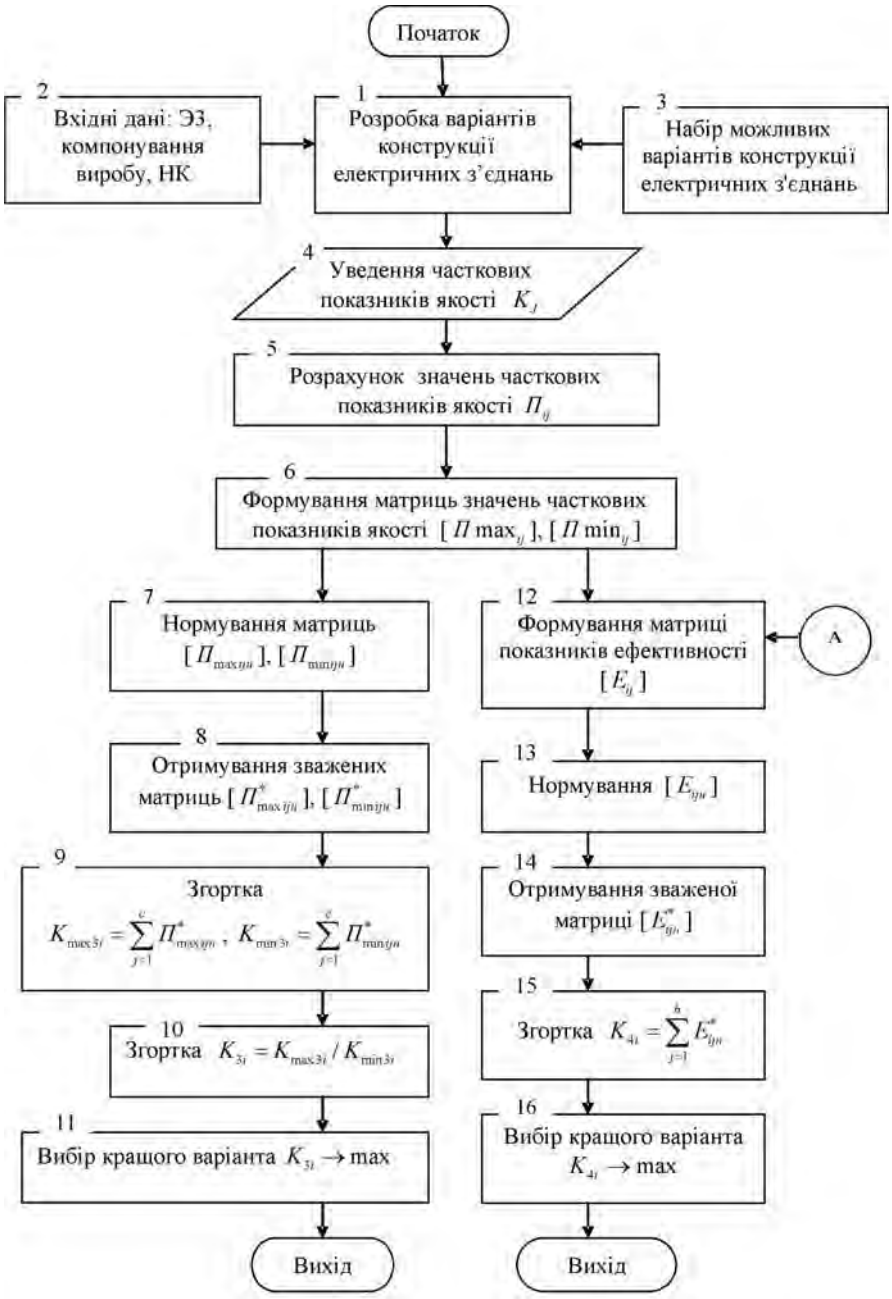


Рис. 3. Алгоритм вибору варіанта конструкції електричних з'єднань за показниками якості та ефективності

терієм трудомісткості, то визначається варіант з мінімальним значенням трудомісткості (**блок 8**). Якщо потрібно врахувати повну вартість електромонтажу, то виконується перехід до блока 9.

Блок 9. Виконується розрахунок вартості трудових операцій виготовлення електромонтажу за прийнятими варіантами. Для цього використовуються отримані в блоці 5 значення трудомісткості, а також трудові нормативи в грошовому виразі (блок 6).

Блок 10. Виконується розрахунок вартості складових частин та матеріалів, що необхідні для створення конструкцій електричних з'єднань. Для цього крім даних, отриманих в блоці 4, використовується база даних, що включає нормативну вартість цих складових частин та матеріалів (**блок 11**).

Блок 12. Виконується розрахунок загальної вартості варіантів конструкції електричних з'єднань. В подальшому отримані значення використовуються для вибору кращого варіанта електромонтажу (**блок 13**), а також для розрахунку показників ефективності в наступному алгоритмі (зв'язок **А**).

2. Алгоритм вибору за показниками якості та ефективності

Алгоритм представлений на рис. 3. Вибір конструкції відбувається на основі порівняння показників якості чи ефективності варіантів конструкції електричних з'єднань. Кращий варіант визначається за максимальними значеннями цих показників.

Блоки 1, 2, 3. Ці блоки такі ж, як і в попередньому алгоритмі.

Блок 4. Уводяться часткові показники якості, що будуть використовуватись для оцінки, та залежності для їх розрахунку. Набір можливих для використання показників якості наведений вище {(4)–(10)}.

Блок 5. Розраховуються значення часткових показників якості за всіма прийнятими до розгляду варіантами конструкції електричних з'єднань.

Блок 6. Формуються дві таблиці з матрицями значень часткових показників якості:

– $[\Pi \max_{ij}]$ – з показниками, що максимізуються K_v, K_m, $t_{пит}$, K_a;

– $[\Pi \min_{ij}]$ – з показниками, що мінімізуються $V_{пит}$, $m_{пит}$, $\tau_{пр}$.

Надалі алгоритм роз'єднується на два напрямки – перший (блоки 7–11) дає змогу вибрати варіант конструкції за показниками якості, другий (блоки 12–16) – за показниками ефективності.

Блок 7. Виконується нормування значень показників якості, використовуючи співвідношення (11). У результаті отримуємо дві пронормовані матриці: $[\Pi \max_{ijн}]$ та $[\Pi \min_{ijн}]$.

Блок 8. Елементи обох матриць помножуються на коефіцієнти вагомості. В результаті отримуємо пронормовані зважені матриці $[\Pi \max^{*}_{ijн}]$ та $[\Pi \min^{*}_{ijн}]$.

Блок 9. З використанням співвідношень (12) виконується згортка значень показників якості для кожного варіанта конструкції електричних з'єднань в межах кожної з двох матриць – за показниками, що максимізуються та мінімізуються.

Блок 10. Виконується згортка загальних значень максимізуємих та мінімізуємих показників якості для кожного варіанта за співвідношенням (13) для визначення оптимального варіанта конструкції (**блок 11**).

Блок 12. Для розрахунку елементів і формування матриці показників ефективності (виграшів) використовуються співвідношення (15).

Блок 13. Нормування значень показників ефективності виконується за співвідношенням (16).

Блок 14. Для отримання пронормованої зваженої матриці $[E^{*}_{ijн}]$ уводяться коефіцієнти вагомості.

Блок 15. Виконується згортка часткових показників ефективності в інтегральний у вигляді суми (17) для кожного з варіантів конструкції електричних з'єднань для вибору оптимального варіанта конструкції (**блок 16**).

Розроблені моделі і алгоритми призначені для практичної реалізації вибору структур (варіантів конструкції) електричних з'єднань та їх оптимізації. Вони можуть використовуватися на різних стадіях проектування, в тому числі і на ранніх, для різних умов розробки і виробництва ЕЗ. Займатися питаннями створення оптимальних конструкцій електричних з'єднань доцільно в тісному поєднанні з вибором оптимальних базових несучих конструкцій [14].

Подальші дослідження та розробки слід розвивати в напрямку створення програмного продукту і баз даних, інших моделей оптимізації та визначення конструктивно-технологічних параметрів різних видів і методів електричних з'єднань.

СПИСОК ЛІТЕРАТУРИ

1. *Ермолович, А.* Оптоэлектрические печатные платы / А. Ермолович // Электронные компоненты и системы. – 2001. – № 12. – С. 3–4.
2. *Фишер, Д.* Реализация оптоэлектронных оснований для печатных плат / Д. Фишер // Печатный монтаж. – 2007. – № 6. – С. 30–32.
3. *Беломытцев, В.* Электромонтаж без отвертки / В. Беломытцев // Современные технологии автоматизации. – 2005. – № 4. – С. 68–71.
4. *Комков, А.* Кристалл – корпус – печатная плата. Проектирование соединений / А. Комков, Г. Хренов // Электроника: наука, технология, бизнес. – 2005. – № 7. – С. 84–86.
5. *Назаров, Е.* Внутренний монтаж функциональных радиоэлектронных блоков / Е. Назаров // Электроника: наука, технология, бизнес. – 2008. – № 3. – С. 36–39.
6. *Лутченков, Л. С.* Аналитический метод определения метрических параметров проводного монтажа / Л. С. Лутченков // Техника средств святи. Сер. Техника проводной связи. – 1986. – Вып. 4. – С. 22–28.
7. *Лутченков, Л. С.* Расчет конструктивних параметров электромонтажа аппаратуры связи / Л. С. Лутченков // Электросвязь. – 1988. – № 11. – С. 53–55.
8. *Ширяев, Ю. Н.* О выборе вида электромонтажа аппаратуры многоканальной святи [Текст] / Ю. Н. Ширяев, А. Э. Бартули // Техника средств связи. Сер. Техника проводной связи. – 1988. – Вып. 2. – С. 45–48.
9. *Ефименко, А. А.* Формализация задачи выбора способа электрического монтажа по критерию стоимости / А. А. Ефименко, Г. К. Яхонтов // Техника средств святи. Сер. Техника проводной связи. – 1988. – Вып. 5. – С. 102–106.
10. *Ефименко, А. А.* Выбор оптимального вида межблочного электромонтажа аппаратуры передачи и обработки информации / А. А. Ефименко, А. Н. Бузин // Средства связи. – 1990. – Вып. 2. – С. 61–65.
11. *Ефименко, А. А.* Формализация задач проектирования межблочных электрических соединений ЭС / А. А. Ефименко, И. Н. Маринов, А. М. Козаревич // Тр. 11-й Междунар. науч.-практич. конф. «СИЭТ-2010». Т. II. – Одесса, 2010. – С. 68.
12. *Ефименко, А. А.* Система показателей качества конструкций межблочных электрических соединений / А. А. Ефименко, А. В. Голов // Технология и конструирование в электронной аппаратуре. – 1998. – № 3–4. – С. 16–18.

13. *Фролов, В. А.* Анализ и оптимизация в прикладных задачах конструирования РЭС : учеб. пособие / В. А. Фролов. – К. : Вища шк., 1991. – 310 с.
14. *Ефименко, А. А.* Оптимальный выбор стандартных несущих конструкций для электронных средств / А. А. Ефименко, А. И. Вильчинский // Технология и конструирование в электронной аппаратуре. – 2010. – № 2. – С. 22–27.

Надійшла 04.11.2010

Ефименко А. А.

ВЫБОР ОПТИМАЛЬНЫХ КОНСТРУКЦИЙ МЕЖБЛОЧНЫХ ЭЛЕКТРИЧЕСКИХ СОЕДИНЕНИЙ ДЛЯ ЭЛЕКТРОННЫХ СРЕДСТВ

Предложены модели и алгоритмы выбора конструкций электрических соединений по критериям стоимости и трудоемкости, показателям качества и эффективности, которые позволяют оптимизировать процесс разработки электронных средств.

Ключевые слова: электрические соединения, межконтактные соединения, оптимизация стоимости, электронные устройства.

Efimenko A. A.

CHOICE OF OPTIMAL INTERBLOCK ELECTRIC CONNECTONS DESIGN FOR ELECTRONIC MEANS

Models and algorithms are proposed for choosing a design of electric contacts by the criteria of cost and labor expenditures as well as quality and efficiency indices, which permit to optimize the electronic means engineering process.

Key words: electric connections, intercontact connections, cost optimization, electronic devices.

УДК 621.314.63

Остренко В. С.

Канд. техн. наук, доцент Запорізької державної інженерної академії

АЛГОРИТМ ВИЗНАЧЕННЯ ПАРАМЕТРІВ ЕКСПОНЕНТ, ЩО АПРОКСИМУЮТЬ ПЕРЕХІДНИЙ ТЕПЛОВИЙ ОПІР ОХОЛОДЖУВАЧА

Запропоновано алгоритм визначення параметрів експонент, що апроксимують графік залежності перехідного теплового опору охолоджувача в часі. Це дає можливість включити охолоджувач у систему розрахунку температури напівпровідникової структури силових напівпровідникових приладів.

Ключові слова: напівпровідниковий прилад, охолоджувач, температура напівпровідникової структури, тепловий опір, параметри експонент, режими охолодження.

Надійність роботи силових напівпровідникових приладів в значній мірі залежить від температури їх напівпровідникової структури. Тому процесам нагріву та охолодження таких приладів завжди приділяється належна увага. Особливо це стосується нестаціонарних режимів навантаження приладів. Температура структури залежить від втрати потужності в приладі та від теплового опору системи «прилад – охолоджувач». Перехідний тепловий опір системи «прилад – охолоджувач» на момент часу t можна визначити за формулою

$$Z_{\text{thja}}(t) = Z_{\text{thjC}}(t) + R_{\text{thCH}} + Z_{\text{thH}}(t), \qquad (1)$$

де $Z_{\text{thjC}}(t)$ – перехідний тепловий опір «структура – корпус приладу»; R_{thCH} – тепловий опір «корпус приладу – контактна поверхня охолоджувача»; $Z_{\text{thH}}(t)$ – перехідний тепловий опір охолоджувача.

Характеристики $Z_{\text{thjC}}(t)$ та параметри R_{thCH} надаються більшістю великих виробників напівпровідникових приладів в інформаційних матеріалах, причому характеристики $Z_{\text{thjC}}(t)$ надаються як у графічному, так і в аналітичному вигляді. В цей же час, характеристики охолоджувачів $Z_{\text{thH}}(t)$ надаються виробниками охолоджувачів у інформаційних матеріалах тільки у графічній формі [1]. Відсутність аналітичної форми представлення перехідного теплового опору охолоджувача значно ускладнює виконання розрахунків теплових режимів роботи напівпровідникових приладів. Тому розробка алгоритму визначення параметрів експонент, що апроксимують графік залежності перехідного теплового опору охолоджувача в часі, є актуальною.

Рекомендується такий алгоритм визначення параметрів експонент, що апроксимують графік залежності перехідного теплового опору охолоджувача та/або силового напівпровідникового приладу.

1. ВИЗНАЧЕННЯ ПОЧАТКОВИХ ДАНИХ ДЛЯ ВИКОНАННЯ РОЗРАХУНКУ

Залежність теплового опору в часі у напівлогарифмічному масштабі показана на рис. 1.

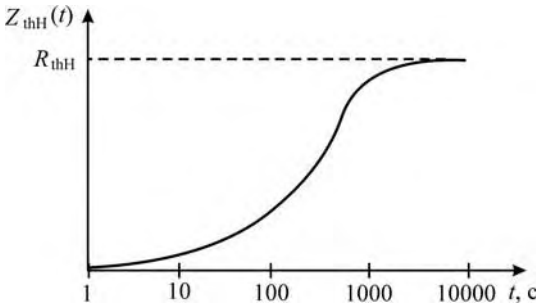


Рис. 1. Залежність перехідного теплового опору в часі

Графік, показаний на рис. 1, можна апроксимувати як суму експонент:

$$Z_{thH}(t) = \sum_{i=1}^{n} R_i[1 - \exp(-t / \tau_i)], \qquad (2)$$

де $Z_{thH}(t)$ – перехідний тепловий опір охолоджувача; R_i, τ_i – параметри експонент, що апроксимують криву перехідного теплового опору, причому сума усіх R_i дорівнює значенню постійного теплового опору, тобто кінцевому значенню $Z_{thH}(t)$; n – кількість експонент, сума яких апроксимує перехідний тепловий опір.

Необхідно визначити: параметри експонент R_i, τ_i на основі графіку рис. 1 та кількість експонент n, яка апроксимує перехідний тепловий опір.

На основі експериментально визначеного графіка залежності перехідного теплового опору в часі, рис. 1, вибираємо точки, що найбільш повно характеризують цю залежність, причому на кожній ділянці часу, кратній 10, повинно бути 2–4 точки, а остання точка повинна бути на ділянці сталого теплового опору R_{thH}.

Координати вибраних точок заносимо у таблицю початкових даних (табл. 1), де $k = 1, 2, 3, \ldots, K$ – порядковий номер точок апроксимації; t_{K-1} – передостання точка в ряду t_k; t_K – остання точка в ряду t_k; $Z_{thH}(t_k)$ – перехідний тепловий опір для моменту часу t_k.

Задаємо похибку апроксимації δ, %.

Таблиця 1. Початкові дані

k	1	2	3	…	$K-1$	K
t_k	t_1	t_2	t_3	…	$t_{(K-1)}$	t_K
$Z_{thH}(t_k)$	$Z_{thH}(t_1)$	$Z_{thH}(t_2)$	$Z_{thH}(t_3)$	…	$Z_{thH}(t_{(K-1)})$	R_{th}

2. ПОРЯДОК ВИКОНАННЯ РОЗРАХУНКУ

2.1. Визначення параметрів експонент, що апроксимують графік перехідного теплового опору, починаємо із заміни графіка нагріву на графік охолодження, як пропонується в роботі [2]. Для цього визначаємо різницю значень теплового опору та перехідного теплового опору

$$Z'_{thH}(t_k) = R_{thH} - Z_{thH}(t_k) \qquad (3)$$

та заносимо дані в табл. 2.

2.2. Визначення параметрів експонент, що апроксимують перехідний тепловий опір, слід починати з моменту часу передостанньої точки $(K-1)$ табл. 2 з координатами $t_k = t_{(K-1)}$; $Z'_{th}(t_k) = Z'_{th}(t_{(K-1)})$, яка стає першою точкою відліку, рис. 2. Другою точкою апроксимації є точка $(K-2)$, яка, згідно з табл. 2, має такі координати: $t_k = t_{(K-2)}$; $Z'_{thH}(t_k) = Z'_{thH}(t_{(K-2)})$.

2.3. Котангенс кута нахилу прямої, що з'єднує точки $(K-1)-(K-2)$, до осі абсцис визначаємо за формулою

$$\operatorname{ctg}\psi_1 = \tau_1 = \frac{t_{(K-1)} - t_{(K-2)}}{\ln Z'_{thH}(t_{(K-2)}) - \ln Z'_{thH}(t_{(K-1)})}, \qquad (4)$$

де τ_1 – перша стала часу першої експоненти у формулі (2).

2.4. Логарифм ординати точки перетину лінії $(K-1)-(K-2)$ з віссю ординат визначаємо за формулою

$$\ln R_1 = \ln Z'_{thH}(t_{(K-1)}) + (t_{(K-1)})/\tau_1, \qquad (5)$$

де R_1 – максимальне значення першої експоненти у формулі (1).

2.5. Значення ординати точки перетину лінії $(K-1)-(K-2)$ з віссю ординат визначаємо потенціюванням $\ln R_1$ за формулою

$$R_1 = \exp(\ln R_1). \qquad (6)$$

Таким чином визначається другий параметр першої експоненти для формули (2).

2.6. Далі перевіряємо, чи належить наступна точка $(K-3)$ лінії $(K-1)-(K-2)$, рис. 2.

Логарифм ординати для моменту часу $t_{(K-3)}$ при τ_1 визначаємо за формулою

$$\ln Z''_{thH}(t_{(K-3)}) = \ln Z'_{thH}(t_{(K-2)}) + (t_{(K-2)} - t_{(K-3)}) / \tau_1. \qquad (7)$$

Таблиця 2. Різниця значень теплового опору та перехідного теплового опору

k	1	2	3	…	$K-2$	$K-1$	K
t_k	t_1	t_2	t_3	…	$t_{(K-2)}$	$t_{(K-1)}$	t_K
$Z'_{thH}(t_k)$	$Z'_{thH}(t_1)$	$Z'_{thH}(t_2)$	$Z'_{thH}(t_3)$	…	$Z'_{thH}(t_{(K-2)})$	$Z'_{thH}(t_{(K-1)})$	0

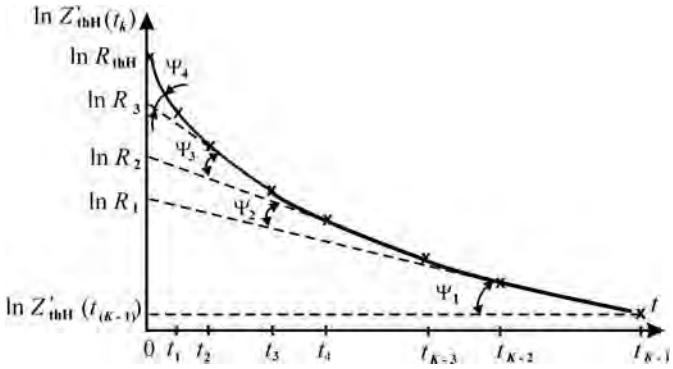


Рис. 2. Залежність логарифму $Z'_{\text{thH}}(t)$ в часі (процес визначення параметрів експонент)

2.7. Значення ординати для моменту часу $t_{(K-3)}$ при τ_1, тобто на лінії $(K-1)-(K-2)$, визначаємо потенціюванням $\ln Z''_{\text{thH}}(t_{(K-3)})$:

$$Z''_{\text{thH}}(t_{(K-3)}) = \exp(\ln Z''_{\text{thH}}(t_{(K-3)})). \qquad (8)$$

2.8. Визначаємо різницю значення $Z''_{\text{thH}}(t_{(K-3)})$ відносно значення $Z'_{\text{thH}}(t_{(K-3)})$ за формулою

$$\delta'' = \left(\frac{Z'_{\text{thH}}(t_{(K-3)}) - Z''_{\text{thH}}(t_{(K-3)})}{Z'_{\text{thH}}(t_{(K-3)})}\right)\cdot 100\%. \qquad (9)$$

2.9. Якщо значення $\delta'' < 0$, тобто воно негативне, це означає, що значення $Z'_{\text{thH}}(t_{(K-3)})$ є хибним, бо не може прилад в режимі охолодження (при відсутності втрат потужності) підвищити свою температуру. В цьому випадку приймаємо, що $Z'_{\text{thH}}(t_{(K-3)}) = Z''_{\text{thH}}(t_{(K-3)})$, тобто точка $(K-3)$ належить експоненті, параметри якої визначені перед цим. Подальший розрахунок слід продовжувати з точки $(K-4)$.

Якщо значення $0 \le \delta'' \le \delta$, де δ – задана точність апроксимації, то це означає, що точка $(K-3)$ належить експоненті, параметри якої визначені перед цим, і подальший розрахунок слід продовжувати з точки $(K-4)$.

Якщо значення $\delta'' > \delta$, це означає, що точка $(K-3)$ не належить експоненті, параметри якої визначені перед цим, і подальший розрахунок слід продовжувати з цієї точки $(K-3)$.

2.10. Визначаємо різницю між значеннями $Z'_{\text{thH}}(t_k)$ (табл. 2) для точок, починаючи з $(k=1)$ до точки, з якої слід продовжувати розрахунок, яка визначена пунктом 2.9 та значеннями експоненти з параметрами (R_i, τ_i), що визначені перед цим, за формулою

$$Z'_{\text{thH}(i+1)}(t_k) = Z'_{\text{thH}}(t_k) - R_i \exp(-t_k/\tau_i), \qquad (10)$$

де $i = 1, 2, \ldots$ – порядковий номер експонент, у яких ще не визначені параметри R_i, τ_i; $k = 1, 2, \ldots, K''$ – індекс часу для точок частини кривої (рис. 2), яка не належить експонентам, у яких вже визначені параметри R_i, τ_i; K'' – індекс часу для останньої точки частини кривої (рис. 2), яка не належить експонентам, у яких вже визначені параметри R_i, τ_i (визначається у пункті 2.9 як точка, з якої слід продовжити розрахунок).

2.11. Наступною точкою відліку призначаємо точку, визначену пунктом 2.9, та продовжуємо розрахунки для визначення параметрів наступної експоненти, повторюючи пункти 2.2–2.10 поки не будуть виконані розрахунками точок 2 – 1.

2.12. Якщо кількість точок значень перехідного теплового опору така, що після виконання розрахунків зі значеннями точок 3 – 2 не урахованим залишається значення точки $(k=1)$ і вона не належить попередній експоненті, то параметри останньої експоненти слід визначати таким чином.

Визначаємо різницю між значенням $Z'_{\text{thH}(n-1)}(t_1)$ та значенням експоненти з визначеними параметрами $(R_{(n-1)}, \tau_{(n-1)})$ за формулою

$$Z'_{\text{thH}n}(t_1) = Z'_{\text{thH}(n-1)}(t_1) - R_{(n-1)} \exp(-t_1/\tau_{(n-1)}). \qquad (11)$$

2.13. Визначаємо параметри останньої експоненти, до якої належить точка $(k=1)$. Оскільки це остання точка, приймаємо її за точку відліку, а за другу точку апроксимації приймаємо значення R_n (значення ординати вище лінії 3 – 2 при $t_0 = 0$), яке визначається за формулою

$$R_n = R_{\text{thH}} - (R_1 + R_2 + \ldots + R_{(n-1)}). \qquad (12)$$

2.14. Котангенс кута нахилу лінії 1 – 0 (прямої, що з'єднує точку 1 та значення $\ln R_n$) до лінії 3 – 2 визначаємо за формулою

$$\operatorname{ctg}\psi_n = \tau_n = \frac{t_1 - 0}{\ln R_n - \ln Z'_{\text{thH}\,n}(t_1)}. \qquad (13)$$

Таким чином завершується визначення параметрів останньої експоненти. Визначено кількість експонент, які апроксимують криву, що зображена на рис. 2; n – кількість експонент, якими апроксимується перехідний тепловий опір та яка визначається внаслідок проведення розрахунків.

2.15. Результатом виконаних розрахунків є параметри експонент, що апроксимують перехідний тепловий опір, які заносяться у табл. 3.

Таблиця 3. Результати розрахунків «назва приладу, охолоджувача, умов охолодження»

i	1	2	3	…	n
R_i, К/Вт	R_1	R_2	R_3	…	R_n
τ_i, с	τ_1	τ_2	τ_3	…	τ_n

2.16. Для визначення похибок у наданні початкових даних та похибок апроксимації необхідно виконати такі розрахунки.

2.16.1. Визначити значення

$$Z_{\text{th Роз}}(t_k) = \sum_{i=1}^{n} R_i[1-\exp(-t_k/\tau_i)], \qquad (14)$$

де $Z_{\text{th Роз}}(t_k)$ – розрахункове значення перехідного теплового опору для моментів часу t_k з табл. 1 (початкові дані) зі значеннями R_i, τ_i за результатами розрахунку, табл. 3.

2.16.2. Визначити абсолютну похибку.

$$\text{Абсолютна похибка} = Z_{\text{th Роз}}(t_k) - Z_{\text{thH}}(t_k), \qquad (15)$$

де $Z_{\text{th Роз}}(t_k)$ – значення, визначені у пункті 2.16.1; $Z_{\text{thH}}(t_k)$ – значення з табл. 1.

2.16.3. Визначити відносну похибку.

$$\text{Відносна похибка} = [(\text{Абсолютна похибка}) / Z_{\text{thH}}(t_k)] \cdot 100\ \%. \qquad (16)$$

Результати розрахунків звести у таблицю з такими стовпчиками: t_k; $Z_{\text{th H}}(t_k)$; $Z_{\text{th Роз}}(t_k)$; Абсолютна похибка; Відносна похибка.

2.17. Назву варіанту, початкові дані, результати розрахунків слід вивести на друк.

3. ПРИКЛАД ВИКОНАННЯ РОЗРАХУНКУ

Визначимо параметри експонент, що апроксимують перехідний тепловий опір охолоджувача типу О253 при швидкості охолоджуючого повітря 6 м/с, який представлено кривою 3 на рис. 3 [3].

Примітка. 0 м/с означає природну конвекцію охолоджуючого повітря, при якій в залежності від розсіюваної потужності швидкість потоку повітря знаходиться в діапазоні значень (0,3–1,2) м/с.

На кривій 3 рис. 3 вибираємо точки, що найбільш повно характеризують залежність перехідного теплового опору в часі при швидкості потоку охолоджуючого повітря 6 м/с, причому за першу точку (k = 1) приймаємо точку для моменту часу t_1 = 2 с, що найбільш точно характеризує початок кривої 3. Параметри вибраних точок заносимо у табл. 4.

Як видно з рис. 3 та з табл. 4, сталим значення перехідного теплового опору стає у точці (k = 8); значення $Z_{\text{thH}}(2000)$ = 0,0975 К/Вт відповідає сталому режиму роботи охолоджувача, тобто R_{thH} = 0,0975 К/Вт. Задаємо, що похибка апроксимації не перевищує δ = 0,5 %.

Визначаємо різницю значень теплового опору та перехідного теплового опору згідно з формулою (3) та заносимо дані в табл. 5.

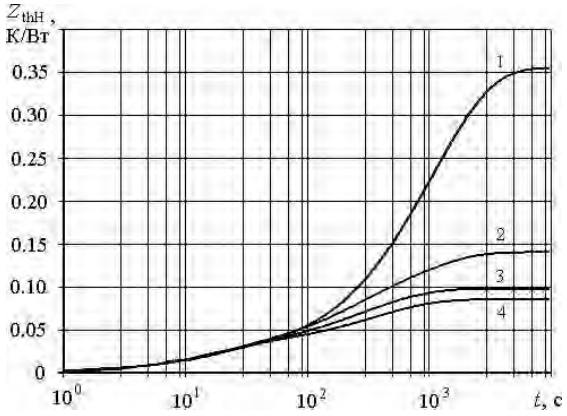


Рис. 3. Перехідний тепловий опір охолоджувача О253 при різних режимах охолодження:
1 – при 0 м/с; 2 – при 3 м/с; 3 – при 6 м/с; 4 – при 12 м/с

Таблиця 4. Значення перехідного теплового опору охолоджувача типу О253 при швидкості охолоджуючого повітря 6 м/с

k	1	2	3	4	5	6	7	8
t_k, с	2	4	10	40	100	400	1000	2000
$Z_{\text{thH}}(t_k)$, К/Вт	0,004	0,0087	0,0161	0,037	0,0485	0,08	0,0928	0,0975

Таблиця 5. Різниця значень теплового опору та перехідного теплового опору охолоджувача типу О253 при швидкості охолоджуючого повітря 6 м/с

k	1	2	3	4	5	6	7	8
t_k, с	2	4	10	40	100	400	1000	2000
$Z'_{\text{thH}}(t_k)$, К/Вт	0,0935	0,0888	0,0814	0,0605	0,049	0,0175	0,0047	0

За першу точку відліку приймаємо точку (k = 7), яка, згідно з табл. 5, має такі координати: t_7 = 1000 с; $Z'_{thH}(t_7)$ = 0,0047 К/Вт. Другою точкою апроксимації є точка (k = 6), яка, згідно з табл. 5, має такі координати: t_6 = 400 с; $Z'_{thH}(t_6)$ = 0,0175 К/Вт.

Котангенс кута нахилу прямої, що з'єднує точки 7 – 6, до осі абсцис визначаємо за формулою (4):

$$\operatorname{ctg}\psi_1 = \tau_1 = \frac{t_7 - t_6}{\ln Z'_{thH}(t_6) - \ln Z'_{thH}(t_7)} =$$
$$= \frac{1000 - 400}{\ln 0{,}0175 - \ln 0{,}0047} =$$
$$= 600/(-4{,}045554 + 5{,}360193) = 456{,}4.$$

Логарифм ординати точки перетину лінії 7 – 6 з віссю ординат визначаємо за формулою (5):

$$\ln R_1 = \ln Z'_{thH}(t_7) + t_7/\tau_1 = \ln 0{,}0047 +$$
$$+ (1000/456{,}4) = -5{,}360193 + 2{,}191 = -3{,}1692.$$

Значення ординати точки перетину лінії 7 – 6 з віссю ординат визначаємо потенціюванням $\ln R_1$ за формулою (6):

$$R_1 = \exp(\ln R_1) = \exp(-3{,}1692) = 0{,}0421 \text{ К/Вт}.$$

Далі перевіряємо, чи належить наступна точка (k = 5) лінії 7 – 6.

Логарифм ординати для моменту часу t_5 = 100 с при τ_1 = 456,4 визначаємо за формулою (7):

$$\ln Z''_{thH}(t_5) = \ln Z'_{thH}(t_6) + (t_6 - t_5)/\tau_1 =$$
$$= \ln 0{,}0175 + (400 - 100)/\ 456{,}4 =$$
$$= -4{,}045554 + 0{,}657318 = -3{,}388236.$$

Значення ординати для моменту часу t_5 = 100 с при τ_1 = 456,4 (тобто на лінії 7 – 6) визначаємо потенціюванням $\ln Z''_{thH}(t_5)$ за формулою (6):

$$Z''_{thH}(t_5) = \exp(\ln Z''_{thH}(t_5)) =$$
$$= \exp(-3{,}388236) = 0{,}03377 \text{ К/Вт}.$$

Визначаємо відносну різницю значення $Z''_{thH}(t_5)$ та значення $Z'_{thH}(t_5)$ за формулою (9):

$$\delta'' = \left(\frac{Z'_{thH}(t_5) - Z''_{thH}(t_5)}{Z'_{thH}(t_5)}\right) \cdot 100\ \% =$$
$$= [(0{,}049 - 0{,}03377)/0{,}049] \cdot 100\ \% =$$
$$= 31\ \% > \delta = 0{,}5\ \%.$$

Це означає, що точка $Z'_{thH}(t_5)$ не належить до цієї експоненти.

Визначаємо різницю між значеннями $Z'_{thH}(t_k)$ (табл. 5) та значеннями експоненти з визначеними параметрами (R_1, τ_1) за формулою (10):

$$Z'_{thH2}(t_k) = Z'_{thH}(t_k) - R_1 \exp(-t_k/\tau_1);$$

$$Z'_{thH2}(t_1) = Z'_{thH}(t_1) - R_1 \exp(-t_1/\tau_1) =$$
$$= 0{,}0935 - 0{,}0421 \cdot \exp(-2/456{,}4) =$$
$$= 0{,}051584 \text{ К/Вт};$$

$$Z'_{thH2}(t_2) = Z'_{thH}(t_2) - R_1 \exp(-t_2/\tau_1) =$$
$$= 0{,}0888 - 0{,}0421 \cdot \exp(-4/456{,}4) =$$
$$= 0{,}04707 \text{ К/Вт};$$

$$Z'_{thH2}(t_3) = Z'_{thH}(t_3) - R_1 \exp(-t_3/\tau_1) =$$
$$= 0{,}0814 - 0{,}0421 \cdot \exp(-10/456{,}4) =$$
$$= 0{,}0402 \text{ К/Вт};$$

$$Z'_{thH2}(t_4) = Z'_{thH}(t_4) - R_1 \exp(-t_4/\tau_1) =$$
$$= 0{,}0605 - 0{,}0421 \cdot \exp(-40/456{,}4) =$$
$$= 0{,}02193 \text{ К/Вт};$$

$$Z'_{thH2}(t_5) = Z'_{thH}(t_5) - R_1 \exp(-t_5/\tau_1) =$$
$$= 0{,}049 - 0{,}0421 \cdot \exp(-100/456{,}4) =$$
$$= 0{,}01518 \text{ К/Вт}.$$

Визначаємо параметри другої експоненти, до якої належать точки (k = 5) та (k = 4).

Тобто, за першу точку відліку приймаємо точку (k = 5), яка має такі координати: t_5 = 100 с; $Z'_{thH2}(t_5)$ = = 0,01518 К/Вт. Другою точкою цієї апроксимації є точка (k = 4), яка має такі координати: t_4 = 40 с; $Z'_{thH2}(t_4)$ = 0,02193 К/Вт.

Котангенс кута нахилу прямої, що з'єднує точки 5 – 4, до лінії 7 – 6 визначаємо за формулою (4):

$$\operatorname{ctg}\psi_2 = \tau_2 = \frac{t_5 - t_4}{\ln Z'_{thH2}(t_4) - \ln Z'_{thH2}(t_5)} =$$
$$= \frac{100 - 40}{\ln 0{,}02193 - \ln 0{,}01518} = 163{,}1.$$

Логарифм ординати точки перетину лінії 5 – 4 з віссю ординат визначаємо за формулою (5):

$$\ln R_2 = \ln Z'_{thH2}(t_5) + t_5/\tau_2 =$$
$$= \ln 0{,}01518 + (100/163{,}1) =$$
$$= -4{,}1877765 + 0{,}613121 = -3{,}5746555.$$

Значення ординати точки перетину лінії 5 – 4 з віссю ординат визначаємо потенціюванням $\ln R_2$ за формулою (6):

$$R_2 = \exp(\ln R_2) = \exp(-3{,}5746555) = 0{,}0280 \text{ К/Вт}.$$

Далі перевіряємо, чи належить наступна точка (k = 3) лінії 5 – 4.

Логарифм ординати для моменту часу t_3 = 10 с при τ_2 = 163,1 визначаємо за формулою (7):

$$\ln Z''_{thH2}(t_3) = \ln Z'_{thH2}(t_4) + (t_4 - t_3)/\tau_2 =$$
$$= \ln 0{,}02193 + (40 - 10)/\ 163{,}1 =$$
$$= -3{,}8198997 + 0{,}183994 = -3{,}635963.$$

Значення ординати для моменту часу $t_3 = 10$ с при $\tau_2 = 163{,}1$ визначаємо потенціюванням $\ln Z''_{thH2}(t_3)$ за формулою (8):

$$Z''_{thH2}(t_3) = \exp(\ln Z''_{thH2}(t_3)) = \exp(-3{,}635963) = 0{,}02636 \text{ К/Вт}.$$

Визначаємо відносну різницю значення $Z''_{thH2}(t_3)$ та значення $Z'_{thH2}(t_3)$ за формулою (9):

$$\delta'' = \left(\frac{Z'_{thH2}(t_3) - Z''_{thH2}(t_3)}{Z'_{thH2}(t_3)}\right)\cdot 100\% = [(0{,}0402 - 0{,}02636)/0{,}0402]\cdot 100\ \% = 34{,}4\ \% > \delta = 0{,}5\ \%.$$

Це означає, що точка $Z'_{thH2}(t_3)$ не належить до другої експоненти, а належить до третьої експоненти.

Визначаємо різницю між значеннями $Z'_{thH2}(t_k)$ та значеннями експоненти з визначеними параметрами (R_2, τ_2) за формулою (10):

$$Z'_{thH3}(t_k) = Z'_{thH2}(t_k) - R_2 \exp(-t_k/\tau_2);$$

$$Z'_{thH3}(t_1) = Z'_{thH2}(t_1) - R_2 \exp(-t_1/\tau_2) = 0{,}051584 - 0{,}028\cdot\exp(-2/163{,}1) = 0{,}02393 \text{ К/Вт};$$

$$Z'_{thH3}(t_2) = Z'_{thH2}(t_2) - R_2 \exp(-t_2/\tau_2) = 0{,}04707 - 0{,}028\cdot\exp(-4/163{,}1) = 0{,}01975 \text{ К/Вт};$$

$$Z'_{thH3}(t_3) = Z'_{thH2}(t_3) - R_2 \exp(-t_3/\tau_2) = 0{,}0402 - 0{,}028\cdot\exp(-10/163{,}1) = 0{,}01386 \text{ К/Вт}.$$

Визначаємо параметри третьої експоненти, до якої належать точки ($k = 3$) та ($k = 2$).

Тобто, за першу точку відліку приймаємо точку ($k = 3$), яка має такі координати: $t_3 = 10$ с; $Z'_{thH3}(t_3) = 0{,}01386$ К/Вт. Другою точкою цієї апроксимації є точка ($k = 2$), яка має такі координати: $t_2 = 4$ с; $Z'_{thH3}(t_2) = 0{,}01975$ К/Вт.

Котангенс кута нахилу прямої, що з'єднує точки 3 – 2, до лінії 5 – 4 визначаємо за формулою (4):

$$\operatorname{ctg}\psi_3 = \tau_3 = \frac{t_3 - t_2}{\ln Z'_{thH3}(t_2) - \ln Z'_{thH3}(t_3)} = \frac{10-4}{\ln 0{,}01975 - \ln 0{,}01386} = 16{,}942.$$

Логарифм ординати точки перетину лінії 3 – 2 з віссю ординат визначаємо за формулою (5):

$$\ln R_3 = \ln Z'_{thH3}(t_3) + t_3/\tau_3 = \ln 0{,}01386 + (10/16{,}942) = -4{,}2787483 + 0{,}59025 = -3{,}6884983.$$

Значення ординати точки перетину лінії 3 – 2 з віссю ординат визначаємо потенціюванням $\ln R_3$ за формулою (6):

$$R_3 = \exp(\ln R_3) = \exp(-3{,}6884983) = 0{,}025 \text{ К/Вт}.$$

Далі перевіряємо, чи належить наступна точка ($k = 1$) до лінії 3 – 2.

Логарифм ординати для моменту часу $t_1 = 2$ с при $\tau_3 = 16{,}942$ визначаємо за формулою (7):

$$\ln Z''_{thH3}(t_1) = \ln Z'_{thH3}(t_2) + (t_2 - t_1)/\tau_3 = \ln 0{,}01975 + (4 - 2)/16{,}942 = -3{,}80655.$$

Значення ординати для моменту часу $t_1 = 2$ с при $\tau_3 = 16{,}942$ визначаємо потенціюванням $\ln Z''_{thH3}(t_1)$ за формулою (8):

$$Z''_{thH3}(t_1) = \exp(\ln Z''_{thH3}(t_1)) = \exp(-3{,}80655) = 0{,}02222 \text{ К/Вт}.$$

Визначаємо відносну різницю значення $Z''_{thH3}(t_1)$ та значення $Z''_{thH3}(t_1)$ за формулою (9):

$$\delta'' = \left(\frac{Z'_{thH3}(t_1) - Z''_{thH3}(t_1)}{Z'_{thH3}(t_1)}\right)\cdot 100\% = [(0{,}02393 - 0{,}02222)/0{,}02393]\cdot 100\ \% = 7{,}1\ \% > \delta = 0{,}5\ \%.$$

Це означає, що точка $Z'_{thH3}(t_1)$ не належить до третьої експоненти, а належить до четвертої експоненти.

Визначаємо різницю між значенням $Z'_{thH3}(t_1)$ та значенням експоненти з визначеними параметрами (R_3, τ_3) за формулою (11):

$$Z'_{thH4}(t_1) = Z'_{thH3}(t_1) - R_3 \exp(-t_1/\tau_3);$$

$$Z'_{thH4}(t_1) = Z'_{thH3}(t_1) - R_3 \exp(-t_1/\tau_3) = 0{,}02393 - 0{,}025\cdot\exp(-2/16{,}942) = 0{,}001714 \text{ К/Вт}.$$

Визначаємо параметри четвертої експоненти, до якої належить точка ($k = 1$). Оскільки це остання точка, приймаємо її за точку відліку, а за другу точку апроксимації приймаємо значення R_4 (значення ординати вище лінії 3 – 2 при $t_0 = 0$), яке визначається за формулою (12):

$$R_4 = R_{thH} - (R_1 + R_2 + R_3) = 0{,}0975 - (0{,}0421 + 0{,}028 + 0{,}025) = 0{,}0024 \text{ К/Вт}.$$

Котангенс кута нахилу лінії 1 – 0 (прямої, що з'єднує точку ($k = 1$) та значення $\ln R_4$) до лінії 3 – 2 визначаємо за формулою (13):

$$\operatorname{ctg}\psi_4 = \tau_4 = \frac{t_1 - 0}{\ln R_4 - \ln Z'_{thH4}(t_1)} = \frac{2-0}{\ln 0{,}0024 - \ln 0{,}001714} = 5{,}941.$$

Результатом виконаних розрахунків є параметри чотирьох експонент ($n = 4$), що апроксимують перехідний тепловий опір охолоджувача типу О253 при швидкості охолоджуючого повітря 6 м/с, які наведені у табл. 6.

Таблиця 6. Результати виконаних розрахунків

i	1	2	3	4
R_i, К/Вт	0,0421	0,028	0,025	0,0024
τ_i, с	456,4	163,1	16,9	5,94

Визначимо похибки апроксимації перехідного теплового опору охолоджувача типу О253 при швидкості охолоджуючого повітря 6 м/с згідно з формулами (14), (15), (16) та наведемо їх у табл. 7.

Таблиця 7. Результати розрахунків похибок апроксимації

t_k, с	$Z_{\text{thH}}(t_k)$, К/Вт	$Z_{\text{th Роз}}(t_k)$, К/Вт	Абсолютна похибка, К/Вт	Відносна похибка, %
2	0,004	0,004	0	0
4	0,0087	0,0075	–0,0012	–13,8
10	0,0161	0,0157	–0,0004	–2,48
40	0,037	0,0347	–0,0023	–6,2
100	0,0485	0,0484	–0,0001	–0,2
400	0,08	0,0776	–0,0024	–3
1000	0,0928	0,0927	–0,0001	–0,1
2000	0,0975	0,0975	0	0

Значні похибки апроксимації можна пояснити недостатньою точністю відліку значень з рис. 3. Для зменшення похибок апроксимації необхідно скорегувати початкові дані у відповідності до табл. 7 та повторити розрахунок.

Таблиця 8. Результати розрахунків параметрів експонент, що апроксимують перехідний тепловий опір охолоджувача О253

Природна конвекція охолоджуючого повітря при потужності розсіювання 220 Вт					
i	1	2	3	4	5
R_i, К/Вт	0,2328	0,05165	0,0491	0,0187	0,00275
τ_i, с	1248,4	1326,4	356,5	13,55	21
Швидкість охолоджуючого повітря 3 м/с					
i	1	2	3	4	
R_i, К/Вт	0,0635	0,05216	0,0219	0,00244	
τ_i, с	865,6	178,6	16,17	6	
Швидкість охолоджуючого повітря 6 м/с					
i	1	2	3	4	
R_i, К/Вт	0,0421	0,028	0,025	0,0024	
τ_i, с	456,4	163,1	16,9	5,94	
Швидкість охолоджуючого повітря 12 м/с					
i	1	2	3	4	
R_i, К/Вт	0,0345	0,0195	0,0286	0,0024	
τ_i, с	496,6	212,4	18,17	5,8	

Результати розрахунків параметрів експонент, що апроксимують перехідний тепловий опір охолоджувача О253 для режимів охолодження, представлених його виробником (рис. 3), наведені в табл. 8.

ВИСНОВОК

Запропонований алгоритм виконання розрахунків для визначення параметрів експонент, що апроксимують перехідний тепловий опір, дозволяє включити охолоджувач в комп'ютерну систему розрахунку температури напівпровідникової структури приладів, що працюють в режимі змінного навантаження.

СПИСОК ЛІТЕРАТУРИ

1. Охладители воздушного охлаждения для приборов таблеточного исполнения [Електронний ресурс] : параметры охладителей / ОАО Электровыпрямитель. – Электрон. дані (1 файл). – Саранск : Электровыпрямитель, 2007. – Режим доступу: http://www.elvpr.ru/poluprovodnikprib/ohladiteli/vozd_tabl.php (вільний). – Назва з екрана.
2. *Давидов, П. Д.* Анализ и расчет тепловых режимов полупроводниковых приборов / Павел Давидович Давидов. – М. : Энергия, 1967. – 144 с.
3. Охладители воздушного охлаждения для приборов таблеточного исполнения типов О343, О253, О353 [Електронний ресурс] : размеры, параметры и характеристики охладителей / ОАО Электровыпрямитель. – Електрон. дані (1 файл). – Саранск : Электровыпрямитель, 2007. – Режим доступу: http://www.elvpr.ru/poluprovodnikprib/ohladiteli/O343%20O253%20O353.pdf (вільний). – Назва з екрана.

Надійшла 08.11.2010

Остренко В. С.

АЛГОРИТМ ОПРЕДЕЛЕНИЯ ПАРАМЕТРОВ ЭКСПОНЕНТ, КОТОРЫЕ АППРОКСИМИРУЮТ ПЕРЕХОДНОЕ ТЕПЛОВОЕ СОПРОТИВЛЕНИЕ ОХЛАДИТЕЛЯ

Предложен алгоритм определения параметров экспонент, аппроксимирующих график зависимости переходного теплового сопротивления охладителя во времени. Это дает возможность включить охладитель в систему расчета температуры полупроводниковой структуры силовых полупроводниковых приборов.

Ключевые слова: полупроводниковый прибор, охладитель, температура полупроводниковой структуры, тепловое сопротивление, параметры экспонент, режимы охлаждения.

Ostrenko V. S.

ALGORITHM FOR DETERMINATION OF EXPONENTS PARAMETERS WHICH APPROXIMATE HEAT SINK TRANSIENT HEAT RESISTANCE

The algorithm is proposed for determination of exponents parameters which approximate heat sink transient heat resistance as a function of time. It permits to include the heat sink into the system for calculation of power semiconductor structure temperature.

Key words: semiconductor device, heat sink, semiconductor structure temperature, heat resistance, exponents parameters, cooling modes.

УДК 681.32

Тімовський А. К.[1], Голдобін О. О.[2]

[1]Канд. техн. наук, доцент Запорізького національного технічного університету
[2]Асистент Запорізького національного технічного університету

ФУНКЦІОНАЛЬНЕ МОДЕЛЮВАННЯ САР ЗА ДОПОМОГОЮ ПРОГРАМИ МАЕС-П

Показано можливість використання програм аналізу електронних схем для функціонального моделювання систем автоматичного регулювання.

Ключові слова: функціональне моделювання, МАЕС-П, система автоматичного регулювання, еквівалентна електрична схема.

ВСТУП

Необхідно відзначити, що об'єкти різної фізичної природи описуються системами звичайних диференційних рівнянь (ЗДР). Це механічні, гідравлічні, теплові системи. Існує аналогія між цими системами, наприклад, аналогами електричної напруги є тиск, температура, швидкість; електричного струму – сили і потоки рідини, газу, теплоти. Існують і аналоги закону Кірхгофа I і II. Це свідчить про те, що інструментарій аналізу електричних схем може бути з успіхом застосований для аналізу систем іншої фізичної природи, особливо на функціональному рівні [1, 2].

Характерним прикладом задач, для функціонального моделювання яких з успіхом можна використовувати програму МАЕС-П [3], є дослідження поведінки систем автоматичного регулювання (САР) або АСУТП, функціональні схеми яких складаються з типових функціональних блоків: диференційних, інтегруючих, нелінійних, підсумовуючих і т. д.

Очевидно, що перетворення функціональних блоків об'єкта регулювання в еквівалентну електричну схему принципових труднощів не викликає. Труднощі виникають при створенні відповідної моделі регулятора. Як правило, регулятор – це складна програма, що реалізує закони управління і регулювання. І бажано, щоб ця програма без змін входила до його моделі на тій мові програмування, на якій написана, тому що відпрацювання цієї програми і є однією із задач моделювання САР.

Тому метою цієї роботи є теоретичне відпрацювання і практична демонстрація можливості використання програми схемотехнічного моделювання МАЕС-П для функціонального моделювання САР за рахунок розробки моделі регулятора у вигляді відповідної нелінійної функції.

МОДЕЛІ ТИПОВИХ ФУНКЦІОНАЛЬНИХ БЛОКІВ САР

Будь який блок (ланку) САР у МАЕС-П можна представити моделлю у вигляді відповідного набору керованих джерел струму, опорів і ємностей, які показують залежність між вихідними і вхідними сигналами цих блоків.

Розглянемо побудову електричних моделей функціональних блоків на прикладі аперіодичного блока.

Рівняння блока: $y = \frac{k}{Tp+1} \cdot x$.

Модель аперіодичного блока у вигляді еквівалентної електричної схеми наведена на рис. 1.

Джерело струму вищенаведеної схеми J_x відображає вхідний сигнал **x** з масштабом $M_x = J_x/x$, а напруга на ємності C – вихідний сигнал **y** з масштабом $M_y = J_y/\mathbf{y}$.

Рівняння цього кола: $U_y = \frac{R}{RCp+1} \cdot J_x$. Підставимо $U_y = M_y \cdot y$, $J_x = M_x \cdot x$ в рівняння моделі і отримаємо $y = \frac{M_x}{M_y} \cdot \frac{R}{RCp+1} \cdot x$.

Із порівняння цього рівняння з рівнянням аперіодичного блока очевидно, що при $M_x = M_y = 1$ $R = k$, $R \cdot C = T$, а $C = T/R$. Значення вхідних і вихідних сигналів моделі і блока будуть однаковими.

Аналогічно розраховуються параметри моделей і інших функціональних блоків, наведених в табл. 1.

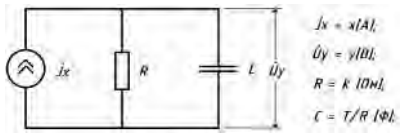
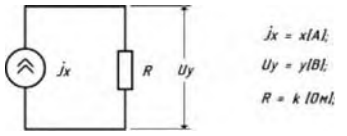
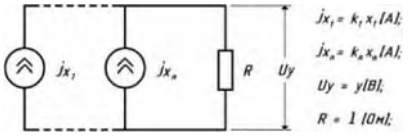
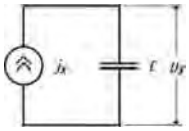
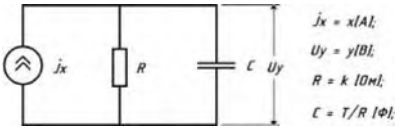
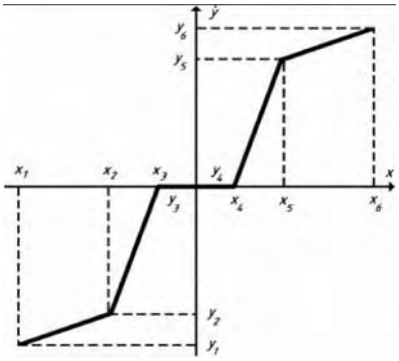
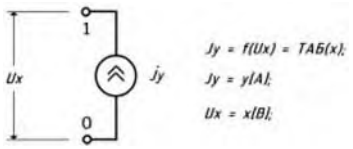


Рис. 1. Еквівалентна електрична схема аперіодичного блока

Таблиця 1. Еквівалентні електричні схеми типових блоків САР

Тип блока	Рівняння блока	Еквівалентна електрична схема
Підсилювач	$y(t) = k \cdot x(t)$	
Суматор	$y(t) = k_1 \cdot x_1(t) + \ldots + k_n \cdot x_n(t)$	
Інтегратор	$T\frac{dy(t)}{dt} = x(t);\ y(t) = \int_0^t \frac{1}{T}x(t)dt;\ y = \frac{1}{PT}x$	
Аперіодичний	$T\frac{dy(t)}{dt} + y(t) = k \cdot X(t);$ $y = \frac{k}{Tp+1}x$	
Нелінійний		 ЭЛЕМЕНТЫ: JY,0-1 = F3(0,0, ТАБХ#U1); ТАБЛИЦЫ: ТАБХ = x1, y1, x2, y2, x3, y3, x4, y4, x5, y5, x6, y6;

ФУНКЦІОНАЛЬНА МОДЕЛЬ РЕАЛЬНОЇ САР

Розглянемо функціональну модель типової САР на прикладі САР палива парового котла АСУ ТП ТЕЦ (рис. 2).

Ця САР підтримує тиск пари в барабані котла та в магістралі на заданому рівні, регулюючи подачу палива. Вона складається з виконавчого органа, об'єкта регулювання (барабана котла та магістралі), датчиків тиску пари в барабані і в магістралі; аналого-цифрових перетворювачів АЦП і регулятора, де UP – подача палива; PR – ознака вмикання/вимикання двигуна ($PR = 1$ – включення двигуна на збільшення подачі палива, $PR = -1$ – на зменшення подачі палива, $PR = 0$ – двигун виключений); $P_б$, $P_н$, $P_м$ – тиск пари в барабані, відбір пари споживачем, тиск пари в магістралі, $I_{рб}$, $I_{рм}$ – струм датчика тиску пари в барабані, струм датчика тиску пари в магістралі; Δ – ціна одного розряду АЦП; $P^ц_б$, $P^ц_{бз}$ – коди поточного і заданого тиску пари в барабані; $P^ц_м$, $P^ц_{мз}$ – коди поточного і заданого тиску пари в магістралі; $PR = f(P^ц_б, P^ц_м)$ – функція регулятора.

Перетворимо функціональну схему САР відповідно до табл. 1 в її еквівалентну електричну схему (рис. 3), яка може бути описана вхідною мовою будь-якої з програм моделювання електронних схем, зокрема вхідною мовою програми МАЕС-П.

Виконавчим органом є клапан подачі палива з електроприводом. Робоча характеристика виконавчого органа відповідає інтегруючому блоку з обмеженням. Рівняння виконавчого органа в еквівалентній електричній схемі САР (рис. 3) моделює інтегруючий блок в складі джерела $J1$ і ємності $C1$, де $J1 = f(E2, E3)$ – джерело струму, що моделює функцію регулятора. Напруга $UC1$ відображає подачу палива UP. Обмеження напруги на ємності $C1$, що відображає обмеження подачі палива, може бути реалізоване за допомогою іншого джерела струму $Jk = f(UC1)$, яке включене паралельно до джерела $J1$ і компенсує його дію при досягненні обмеження.

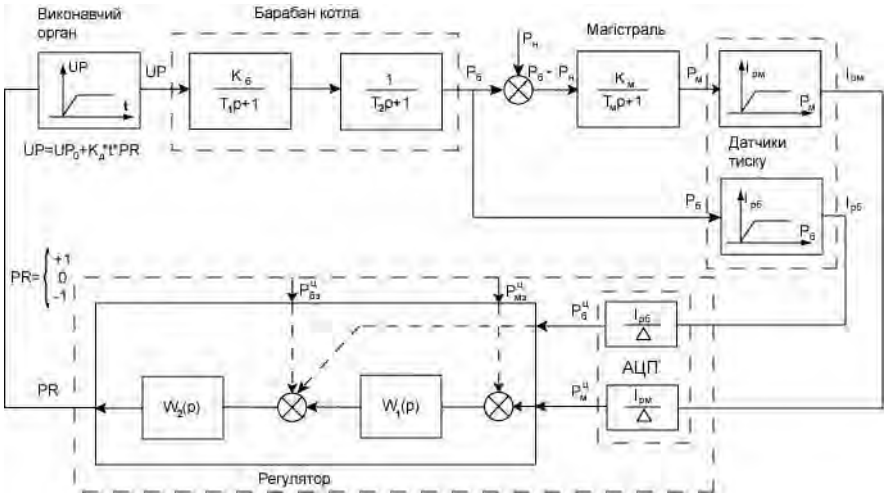


Рис. 2. Функціональна схема САР палива

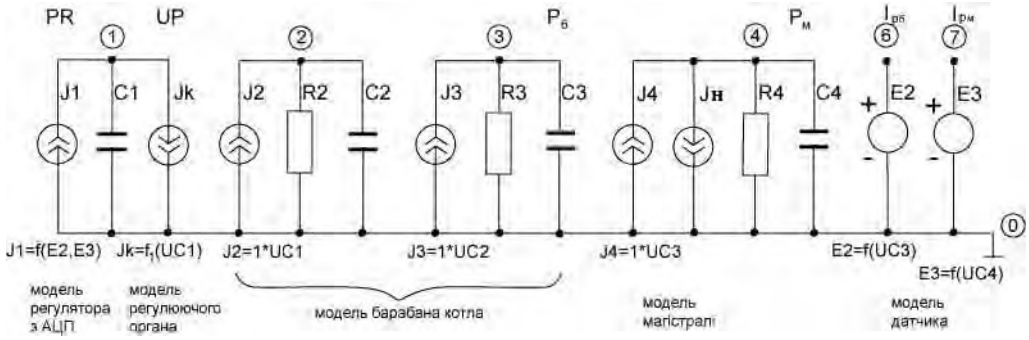


Рис. 3. Еквівалентна електрична схема функціональної моделі САР

Модель барабана котла – це два аперіодичні блоки. На рис. 3 вони відображені елементами *J*2, *R*2, *C*2 і *J*3, *R*3, *C*3, де джерело струму $J2 = 1 \cdot UC1$ повторює вхідний сигнал *UP*, джерело струму $J3 = 1 \cdot UC2$ повторює величину *UC*2, а напруга *UC*3 відображає тиск пари $P_б$ в барабані.

Модель магістралі – це аперіодичний блок, у якого вхідний сигнал має дві складові: сигнал $P_б$ та навантаження $P_н$. Навантаження $P_н$ (зміна тиску пари за рахунок відбору пари споживачами) задається таблицею $P_н = f(t)$. На еквівалентній схемі САР (рис. 3) це відповідні джерела струму $J4 = 1 \cdot UC3$ і $J_н$, а також конденсатор *C*4, напруга на якому відображає тиск пари $P_м$ в магістралі.

Моделі датчиків тиску $P_б$ і $P_м$ – це підсилюючі блоки з обмеженнями, які на рис. 3 зображені джерелами напруги $E2 = f(UC3)$ і $E3 = f(UC4)$, що відображають відповідні характеристики датчиків тиску $I_{рб} = f(P_б)$, $I_{рм} = f(P_м)$.

Модель регулятора представлена на рис. 3 нелінійним джерелом струму $J1 = f(E2, E3)$, що описується спеціальною нелінійною функцією, в яку входять програма контролера ($PR = f(P^ц_б, P^ц_м)$) і моделі АЦП, що перетворюють аналогові сигнали $P_б$ та $P_м$ в цифрову форму для контролера-регулятора.

МОДЕЛЬ РЕГУЛЯТОРА

Програма МАЕС-П має спеціальний механізм введення нових нелінійних функцій самим користувачем, чого не мають інші аналогічні програми. Це дозволяє ввести програму регулятора як нелінійну функцію в МАЕС-П. Якщо контролер регулятора сумісний з IBM PC, то його програма переноситься в нелінійну функцію без змін.

Програма МАЕС-П написана мовою FORTRAN-77 і, якщо програма контролера написана іншою мовою, наприклад C, то нелінійна функція повинна складатися з оболонки, яка написана на FORTRAN-77, для об'єднання програми МАЕС-П з програмою контролера на мові C без змін у вигляді підпрограми.

В оболонці реалізовано також модель аналого-цифрового перетворювача (АЦП), алгоритм циклічної видачі управляючих імпульсів з заданим робочим тактом і узгодження з ним змінного кроку інтегрування, особливо в разі його відкидання при незбіжності ітерацій Ньютона. Цю нелінійну функцію F59 розроблено відповідно до правил розробки нових нелінійних функцій для програми МАЕС-П.

Блок-схема алгоритму нелінійної функції регулятора з АЦП наведена на рис. 4.

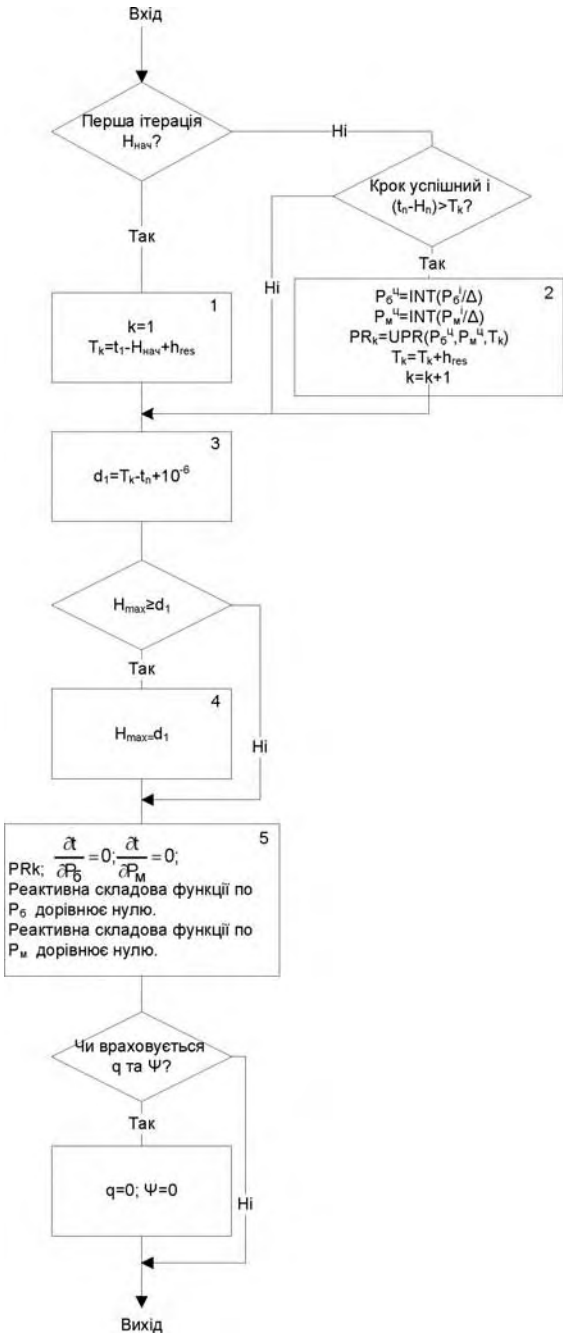


Рис. 4. Блок-схема алгоритму нелінійної функції регулятора з АЦП

На вхід нелінійної функції як аргументи надходять напруги джерел $E2$ і $E3$, що відповідають показанням датчиків тиску $I_{рб} = f(P_б)$ і $I_{рм} = f(P_м)$. В оболонці нелінійної функції вони перетворюються з аналогових величин в цифрові $P_б^ц$, $P_м^ц$ і надходять до підпрограми регулятора *UPR* ($P_б^ц$, $P_м^ц$, h_{res}), де порівнюються з заданими значеннями і формується управляючий сигнал *PR*. Глобальними змінними є t_n – поточний час інтегрування, h_n – поточний крок інтегрування, параметри h_{res} – крок управління, $T_k = k \cdot h_{res}$ – кінцевий час k-го кроку управління.

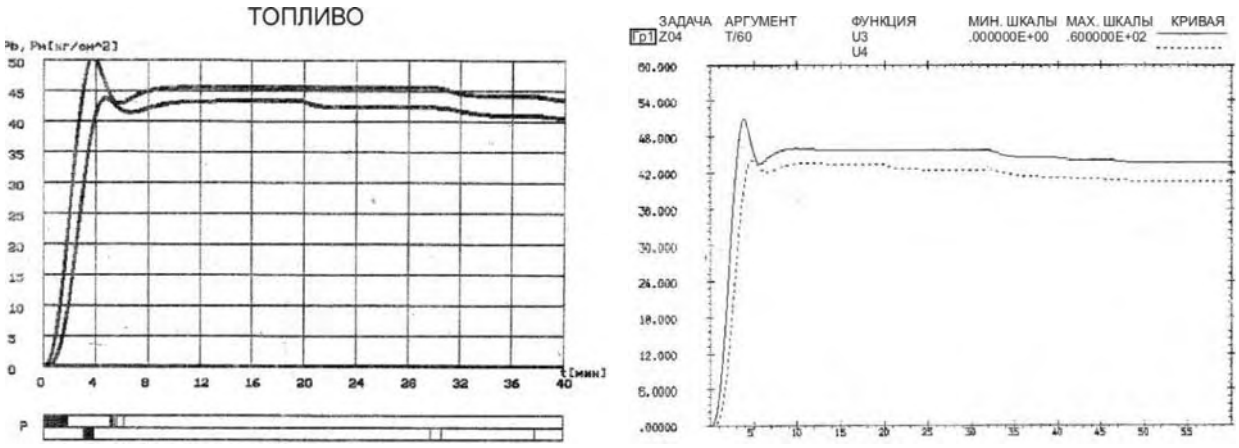
Виходом функції є значення всіх складових нелінійної функції, які беруть участь у формуванні матриці провідностей і правої частини: повне і безреактивне значення функції і частинні похідні вихідної змінної (реактивні і безреактивні) по кожному з аргументів функції. Але оскільки виходом функції є імпульсний сигнал без будь-яких затримок, то всі ці складові дорівнюють нулю, за винятком, безумовно, значення самої функції, яке виробляє підпрограма регулятора UPR.

Якщо розрахунок тільки починається, тобто виконується перша ітерація початкового кроку інтегрування $h_n = h_{нач}$, $t_n = t_{нач}$, то, щоб розрахувати кінцевий час першого кроку управління $T1$, необхідно повернутися в точку $t = 0$, тобто $T1 = (t_n - h_{нач}) + h_{res}$.

На цьому кроці управління блок 2 з моделями АЦП і підпрограмою регулятора не підключається, тому що $PR = 0$ (початкове значення), процес інтегрування йде з кроком h_n, що автоматично вибирається і обмежується блоком 4 до значення $d1$, що дорівнює відрізку часу від точки $t = t_n$ до точки $t = T1$, для того, щоб останній успішний крок інтегрування попав в точку $t = T1 + D$ ($D = 0{,}000001$ для гарантії, що точку $T1$ пройдено).

На наступних кроках управління, якщо останній був успішним на k-му кроці управління, крок інтегрування h_{n-1} попав в точку $t = T_k + 0{,}000001$ і вибрано новий крок h_n, тобто $t_{n-1} + h_n > T_k$, то в точці $t = t_n - h_n = T_k + 0,000001$ підключається блок 2 з моделями АЦП і підпрограмою регулятора, який виробляє черговий імпульс управління PR_k і встановлює черговий крок управління, кінцевий час якого $T_k = T_k + h_{res}$.

Вихідний імпульс *PR* залишається незмінним на протязі всього кроку управління, поки процес інтегрування не дійде кінця цього кроку, і тоді знову підключається підпрограма регулятора і формується новий управляючий імпульс, і так далі.



а) *б*)

Рис. 5. Перехідний процес встановлення тиску в барабані котла $P_б$ та в магістралі $P_м$:
а – копія екрана спеціалізованої програми АСУТП; *б* – результати функціонального моделювання

РЕЗУЛЬТАТИ МОДЕЛЮВАННЯ

Авторами було проведено моделювання перехідного процесу встановлення тиску пари в котлі реальної САР палива в АСУТП Коровинецького цукрового заводу на спеціалізованій програмі функціонального моделювання САР (рис. 5, *а*) і на програмі МАЕС-П (рис. 5, *б*).

Результати моделювання повністю співпадають.

ВИСНОВКИ

В статті наведені електричні схеми моделей типових функціональних блоків регулювання і розроблена еквівалентна електронна схема функціональної моделі реальної САР. Показана особливість настроювання і керування роботою регулятора у складі функціональної моделі САР. Наведений приклад доказує адекватність такого підходу до функціонального моделювання.

Набір даних, що був отриманий під час роботи спеціальної програми САР палива, характеризує перехідний процес встановлення тиску в барабані котла. Ці дані співпадають з результатами, які дає сеанс функціонального моделювання за допомогою програми МАЕС-П.

Період часу функціонального моделювання значно менший за час спрацьовування регулятора реальної АСУТП. Це підтверджує практичну цінність виконаної роботи. Випереджаючий характер функціонального моделювання може значно спростити процес прийняття рішення з попереднім передбаченням наслідків.

Таким чином, показано можливість використання програми МАЕС-П для функціонального моделювання САР. Це дозволяє при наявності програми схемотехнічного моделювання відмовитись від придбання спеціалізованої програми функціонального моделювання.

СПИСОК ЛІТЕРАТУРИ

1. *Норенков, И. П.* Введение в автоматизированное проектирование технических устройств и систем / И. П. Норенков. – М. : Высш. школа, 1986. – 304 с.
2. *Петренко, А. И.* Автоматизация схемотехнического проектирования в машиностроении / Петренко А. И., Ладогубец В. В., Чкалов В. В. – Киев : УМК ВО, 1988. – 180 с.
3. *Піза, Д. М.* Моделювання радіоелектронних пристроїв : навчальний посібник / Піза Д. М., Тімовський А. К., Лугін А. І. – Запоріжжя : ЗНТУ, 2003. – 258 с.

Надійшла 13.05.2010
Після доробки 19.11.2010

Тимовский А. К., Голдобин А. А.

ФУНКЦИОНАЛЬНОЕ МОДЕЛИРОВАНИЕ САР С ПОМОЩЬЮ ПРОГРАММЫ МАЕС-П

Показана возможность использования программ анализа электронных схем для функционального моделирования систем автоматического регулирования.

Ключевые слова: функциональное моделирование, МАЕС-П, система автоматического регулирования, эквивалентная электрическая схема.

Timovsky A. K., Goldobin A. A.

FUNCTIONAL MODELING OF ACS USING THE MAEC-П PROGRAM

Functional modeling of automatic control systems using the programs of electronic circuits analysis.

Key words: functional modeling, МАЕС-П, automatic control system, equivalent electric circuit.

МАТЕМАТИЧНЕ ТА КОМП'ЮТЕРНЕ МОДЕЛЮВАННЯ

МАТЕМАТИЧЕСКОЕ И КОМПЬЮТЕРНОЕ МОДЕЛИРОВАНИЕ

MATHEMATICAL AND COMPUTER MODELLING

УДК 004.056.53

Коломыцев М. В.[1], Носок С. А.[1]

[1]Канд. техн. наук, доцент Национального технического университета Украины «Киевский политехнический институт»

УЯЗВИМОСТИ ПРИЛОЖЕНИЙ К НЕКОРРЕКТНЫМ ВХОДНЫМ ДАННЫМ

В данной статье рассматриваются различного рода уязвимости, присущие приложениям без достаточного контроля входных данных. Приводятся характерные признаки таких уязвимостей и даются рекомендации по их устранению.

Ключевые слова: уязвимость, приложения, корректные данные, некорректные данные.

ВВЕДЕНИЕ

Одной из основных причин уязвимостей программного обеспечения является неполная проверка входных данных либо ее полное отсутствие. Входные данные могут поступать от пользователей, из хранилища данных, сетевых сокетов и других источников. Любые данные, поступающие в приложение извне, перед их использованием должны проверяться для определения их корректности. Под корректностью данных в этом случае понимается их соответствие требованиям приложения – к формату, диапазону допустимых значений и т. п.

В процессе контроля приложение должно проверять, относятся ли данные к заведомо корректным, и отвергать их, если убедится, что это не так. Нельзя строить контроль входных данных на проверке того, относятся ли полученные данные к заведомо некорректным или потенциально опасным, поскольку атакующий может использовать различные приемы для того, чтобы обойти процедуру такой проверки. Он может по-иному представить фрагменты входных данных, либо использовать структуры данных, отсутствующие в образцах потенциально опасных.

Тот факт, что обрабатываться будут только заведомо корректные данные, может означать, что часть поступившей корректной информации будет отвергнута, как не прошедшая проверку, однако только так можно существенно снизить вероятность обработки некорректных данных.

Таким образом, разработчик обязан заложить в приложение проверку всех поступивших данных, как пользовательского ввода, так и данных, поступивших из других источников.

Анализу уязвимостей приложений вообще и рассматриваемого класса уязвимостей в частности посвящено много работ и информационных источников [1, 2].

Рассмотрим основные виды уязвимостей и способы их устранения.

ПОСТАНОВКА ЗАДАЧИ

Безопасность приложения предполагает, что данные, обрабатываемые, передаваемые и хранимые с помощью приложения, защищены от неавторизованного доступа. Угроза неавторизованного доступа возникает, если в приложении присутствуют определенные уязвимости. В данной статье ставится задача определения перечня уязвимостей приложений, обусловленных недостаточно полной проверкой данных, поступивших в приложение извне. Определяются характерные признаки, свидетельствующие о наличии таких уязвимостей, и предлагаются меры, направленные на их устранение.

ВНЕДРЕНИЕ КОДА В SQL ЗАПРОСЫ (SQL INJECTION)

Данная уязвимость позволяет модифицировать запросы к базе данных, с тем, чтобы получить несанкционированный доступ к данным. Данная уязвимость эксплуатируется через недостаточную проверку входных данных.

Следующие особенности кода приложения свидетельствуют о потенциальной возможности SQL injection:

– вводимые пользователем данные без проверки используются для построения запросов или как параметры хранимых процедур;

– при построении запросов используется конкатенация строк;

– при построении запросов используется замена строк;

– для выполнения запроса используется функция exec (в случае использования СУБД Microsoft SQL Server).

Основным методом обнаружения данной уязвимости является анализ (ревизия) исходных текстов программного обеспечения. Принципиально важно, чтобы все приложения, работающие с базами данных, были проанализированы на предмет наличия данной уязвимости.

Чтобы минимизировать вероятность возникновения данной уязвимости, следует придерживаться следующих правил:

– для проверки текстов программ использовать инструментальные средства анализа, обладающие низкой вероятностью ошибок 2-го рода для SQL injection (т. е. низкой вероятностью пропуска уязвимости в исходных кодах приложения). Существует широкий набор инструментальных средств для анализа исходного кода [3]. Например, Ounce Labs, Parasoft's Jtest & C++test, Klocwork Insight, Fortify Source Code Analyzer (SCA), GrammaTech CodeSonar, Coverity Prevent;

– проверять все данные, введенные пользователем, разрешая только заведомо корректные данные;

– использовать параметризированные запросы;

– не использовать конкатенацию или замену строк при построении кода запроса в тексте приложения;

– не использовать учетную запись администратора для подключения к БД;

– для доступа к данным использовать представления, непосредственный доступ к таблицам должен быть запрещен;

– для выполнения динамически созданного запроса использовать команду sp_executesql вместо exec (для СУБД Microsoft SQL Server);

– для проверки параметров хранимых процедур использовать функцию quotename (для СУБД Microsoft SQL Server).

УЯЗВИМОСТЬ ЦЕЛОЧИСЛЕННОГО ПЕРЕПОЛНЕНИЯ

Уязвимость целочисленного переполнения возникает как результат некорректных операций над данными целого типа [4]. Последствия целочисленного переполнения могут быть самыми разными – от некорректного результата до краха приложения. Данная уязвимость может быть использована для изменения значения критически важных данных – изменения размера буфера, значения индекса массива. При этом возникает возможность переполнения буфера. К возникновению данной уязвимости могут привести следующие операции:

– совместное оперирование знаковыми и беззнаковыми целыми (например, сравнение знакового целого и беззнакового);

– усечение целых (например, усечение 32-битового целого до 16-битового);

– потеря значимости и переполнение (например, в результате суммирования двух целых может быть получено число большее, чем максимально возможное для целого типа данных).

Следующие особенности кода приложения свидетельствуют о потенциальной возможности уязвимости:

– смешивание знаковых и беззнаковых целых в операциях вычисления и сравнения;

– смешивание данных различных типов в операциях вычисления и сравнения;

– сравнение переменных и литералов;

– отсутствие проверки входных данных;

– использование результата вычисления без его проверки.

Основным методом обнаружения данной уязвимости является ревизия исходных текстов программного обеспечения. Важно, чтобы все случаи динамически выделяемой памяти и индексных массивов, использующих целочисленную арифметику, были проанализированы на предмет корректности. В процессе такой проверки необходимо тестировать следующие ситуации:

– ввод отрицательных значений при запросе на ввод целых чисел;

– ввод целых, соответствующих граничным значениям хранения данных в одном байте, двух байтах и т. д. – т. е. чисел 0, 7, 8, 254, 255, 16353, 16354;

– ввод очень длинных строк (более 64 К);

– ввод строк, длина которых равна типичным граничным значениям (32К, 32К-1, 64К-1, 64К);

– ввод случайных, непредусмотренных или неверных данных – так называемый Fuzz testing. Fuzz testing – это техника тестирования, состоящая в подаче на вход приложения случайных и направленно сформированных наборов данных с целью генерации ошибок в приложении или его аварийного завершения. Процентный показатель сбоев и крахов приложения является показателем уязвимости.

Чтобы минимизировать вероятность возникновения данной уязвимости, следует придерживаться следующих правил:

– для проверки текстов программ использовать инструментальные средства анализа, обладающие низкой вероятностью ошибок 2-го рода для уязвимостей данного типа;

– везде, где возможно, использовать беззнаковые целые;

– при выделении памяти использовать только беззнаковые целые;

– при построении индексированных массивов использовать только беззнаковые целые;

– проверять введенные пользователем числовые данные, разрешая только заведомо корректные данные;

– при компиляции приложения устанавливать максимально подробный уровень сообщений компилятора.

УЯЗВИМОСТИ СТРОК ФОРМАТИРОВАНИЯ (FORMAT STRING)

Наличие данной уязвимости позволяет передать в качестве входного параметра функции ввода/вывода специальным образом сконструированную строку, что позволяет атакующему получить информацию об управлении приложением и даже изменить ход выполнения приложения [4]. Многие функции ввода/вывода позволяют нужным образом отформатировать строку, переданную им в качестве входного параметра. Это означает, что входные параметры функции могут содержать специальные символы, определяющие формат преобразования. Специальным образом подобранные символы форматирования позволяют прочесть содержимое областей памяти, организовать переполнение буфера. В худшем случае эксплуатация уязвимости позволяет атакующему выполнить произвольный код в системе. Первопричиной такой уязвимости является недостаточная проверка входных данных.

Основным методом обнаружения уязвимости строк форматирования является анализ исходных текстов программ. Следует очень осторожно использовать такой спецификатор форматирования, как %p (указатель) и избегать спецификатора %n (количество символов, записанных по адресу, указанному в качестве второго аргумента).

При тестировании приложения следует вставлять спецификаторы форматирования во вводимые данные во всех точках ввода строковых данных и анализировать полученный результат. Количество и вид спецификаторов существенно варьируется в зависимости от используемого языка программирования. При тестировании, кроме спецификаторов, определенных в используемом языке, следует обязательно проверять спецификаторы, определенные в языках C/C++.

Чтобы минимизировать вероятность появления данной уязвимости, необходимо придерживаться следующих правил:

– для проверки текстов программ использовать инструментальные средства анализа, обладающие низкой вероятностью ошибок 2-го рода для уязвимостей данного типа;

– проверять все полученные извне данные перед обработкой, разрешая только заведомо корректные данные;

– функции форматирования строк, используемые в приложении, должны быть доступны только привилегированным пользователям;

– если исходный код приложения написан на языке C++, нужно использовать операторы управления потоком (stream operators) вместо функций семейства printf;

– если используется компилятор GCC, необходимо устанавливать режимы -Wformat, -Wformat-security для обнаружения ошибок использования функций форматирования строк;

– предпочтение следует отдавать компилятору Microsoft Visual C++ 2005 или более позднему для обнаружения ошибок использования функций форматирования строк в процессе выполнения.

УЯЗВИМОСТЬ ВНЕДРЕНИЯ КОМАНД (COMMAND INJECTION)

При эксплуатации данной уязвимости в приложение могут быть переданы данные, приводящие к тому, что у атакующего появляется возможность манипулировать командной оболочкой (shell) операционной системы. В результате могут быть выполнены команды с административными полномочиями.

Потенциально, уязвимость возникает, если в приложении используются команды порождения процессов, параметры которых не проверяются должным образом. Список таких команд для разных языков программирования приведен в табл. 1.

Таблица 1. Команды порождения процессов для разных языков программирования

Язык программирования	Потенциально опасные функции
C/C++	system (), popen (), execlp (), execvp (), ShellExecute (), ShellExecuteEx (), _wsystem ()
Perl	system, exec, `, open, |, eval, /e
Python	exec, eval, os.system, os.popen, execfile, input, compile
Java	Class.forName (), Class.newInstance (), Runtime.exec ()

В дополнение к ревизии исходных текстов программ необходимо провести тестирование приложения на предмет обнаружения уязвимости внедрения команд. В процессе тестирования необходимо:

– идентифицировать все интерпретаторы и компиляторы, используемые для создания и функционирования приложения;

– идентифицировать все потенциально опасные символьные последовательности, которые могут изменить характер действий интерпретатора;

– сформировать входные данные, содержащие такие потенциально опасные наборы символов и проанализировать видимые последствия обработки таких данных.

УЯЗВИМОСТЬ К МЕЖСАЙТОВОМУ СКРИПТИНГУ (CROSS SITE SCRIPTING – XSS)

Уязвимость XSS возникает, когда входные данные, получаемые Web-сервером от одного клиента, могут быть пересланы другому клиентскому браузеру через Web страницу [5]. Эти данные могут содержать программный код, например на JavaScript, который будет выполнен браузером клиента. Поскольку система безопасности клиента считает, что код получен от Web-сервера, то злоумышленный код может получить доступ к данным, доступ к которым имеет Web-сервер (например, файлам cookie), либо изменить Web страницу, манипулируя ссылками на ней и т. п. Эксплуатируя данную уязвимость, атакующий может получить доступ к персональным данным пользователя либо перенаправить пользователя на нужный атакующему сайт.

Если пользовательский ввод возвращается в браузер, то приложение потенциально уязвимо для XSS атак.

В дополнение к ревизии исходных текстов программ необходимо провести тестирование приложения на предмет обнаружения уязвимости к XSS атакам. В процессе тестирования необходимо:

– выполнять запрос к приложению, помещая в пользовательский ввод заведомо неправильные данные;

– проанализировать ответный HTML код с целью обнаружения в нем отправленных данных. Причем необходимо учитывать, что отправленные данные могут быть возвращены не сразу, а в ответ на последующие запросы.

Для минимизации вероятности возникновения XSS уязвимости следует:

– для проверки текстов программ использовать инструментальные средства анализа, обладающие низкой вероятностью ошибок 2-го рода для уязвимостей данного типа;

– контролировать все входные данные, разрешая только заведомо правильные данные;

– если во входных данных необходимо использовать специальные символы HTML-кода, необходимо удалять все HTML-теги, кроме разрешенных;

– устанавливать для Web страниц определенную кодовую страницу, чтобы избежать ввода непредусмотренных данных.

УЯЗВИМОСТЬ К ПЕРЕПОЛНЕНИЮ БУФЕРА (BUFFER OVERFLOW)

Данная уязвимость возникает, если существует возможность записи в буфер информации больше, чем позволяет область памяти, выделенная под буфер. Существует несколько вариантов переполнения буфера [4], все они потенциально опасны для приложения и могут привести к аварийному завершению приложения либо к выполнению злоумышленного

кода в атакуемой системе. Если приложение выполняется от имени привилегированной учетной записи, такой как system или root, последствия для атакуемой системы могут быть катастрофическими. Эксплуатация данной уязвимости в основном осуществляется при недостаточной проверке входных данных.

Следующие особенности приложения могут свидетельствовать о наличии уязвимости переполнения буфера:

– входные данные не проверяются перед тем, как будут скопированы в буфер;

– некорректное использование небезопасных функций;

– некорректное определение необходимых размеров буфера;

– некорректное вычисление размера индексных массивов.

Основным методом обнаружения уязвимости является анализ исходных текстов программ и fuzz testing. Чтобы минимизировать вероятность возникновения данной уязвимости, следует придерживаться следующих правил:

– для проверки текстов программ использовать инструментальные средства анализа, обладающие низкой вероятностью ошибок 2-го рода для уязвимостей данного типа;

– проверять все полученные извне данные перед обработкой, разрешая только заведомо корректные данные;

– не использовать заведомо небезопасные функции (такие как strcpy, strcat, lstrcat, sprintf, fprintf и другие), заменяя их безопасными;

– тщательно проверить все алгоритмы вычисления размера буфера;

– при компиляции использовать опции компилятора для контроля переполнения буфера. Например, в компиляторе Visual C++ 2005 Service Pack 1 и более поздних следует использовать опции /GS, /SAFESEH, /NXCOMPAT и /DYNAMICBASE. В компиляторе gcc 4.1.2-25 и более поздних используйте опцию -stack-protector.

ВЫВОДЫ

Широкое распространение распределенных информационных систем, Internet-технологий существенно увеличило множество точек доступа к приложениям и, как следствие, расширило возможности злоумышленников атаковать информационные системы. Большинство таких атак становится возможным в силу наличия уязвимостей в приложениях, обрабатывающих данные из внешних источников. К сожалению, традиционные методы защиты периметра (firewalls, email filters и пр.) не способны полностью решить задачу защиты от внешних атак. Разработчики приложений должны знать характерные признаки уязвимостей и уметь таким образом строить код приложения, чтобы исключить возможность атак, эксплуатирующих такие уязвимости. В статье рассматриваются популярные виды уязвимостей, описываются их признаки и даются рекомендации по устранению уязвимостей.

СПИСОК ЛИТЕРАТУРЫ

1. National Vulnerability Database [Электронный ресурс] / National Institute of Standards and Technology. – Режим доступа: http://nvd.nist.gov/, свободный. – Загл. с экрана.
2. List of Common Vulnerabilities and Exposures [Электронный ресурс] / The MITRE Corporation. – Режим доступа: http://cve.mitre.org/, свободный. – Загл. с экрана.
3. Поиск уязвимостей в программах с помощью анализаторов кода [Электронный ресурс] / Елена Харитонова. – Режим доступа: http://www.codenet.ru/progr/other/code-analysers.php, свободный. – Загл. с экрана.
4. *Thompson, H. H.* The Software Vulnerability Guide (Programming Series) / Herbert H. Thompson, Scott G. Chase. – Pontypridd, UK, United Kingdom : Charles River Media, 2005. – 354 p.
5. *Хогланд, Г.* Взлом программного обеспечения: анализ и использование кода / Грег Хогланд, Гари Мак-Гроу. – М. : Вильямс, 2005. – 389 с. : ил.

Надійшла 4.11.2010

Коломицев М. В., Носок С. О.

ВРАЗЛИВОСТІ ДОДАТКІВ ДО НЕКОРЕКТНИХ ВХІДНИХ ДАНИХ

У даній статті розглядаються різного роду вразливості, властиві додаткам без достатнього контролю вхідних даних. Наводяться характерні ознаки таких вразливостей і даються рекомендації з їхнього усунення.

Ключові слова: вразливість, додатки, коректні дані, некоректні дані.

Kolomytsev M., Nosok S.

APPLICATIONS VULNERABILITIES CAUSED BY INCORRECT INPUT DATA

Different kinds of vulnerabilities typical for applications without sufficient input data control are discussed in this article. Typical characteristics of such vulnerabilities are given and elimination recommendations are proposed.

Key words: vulnerability, applications, correct data, incorrect data.

УДК 519.816

Колпакова Т. А.

Магистр Запорожского национального технического университета

ОПРЕДЕЛЕНИЕ КОМПЕТЕНТНОСТИ ЭКСПЕРТОВ ПРИ ПРИНЯТИИ ГРУППОВЫХ РЕШЕНИЙ

Рассматривается проблема учета компетентности экспертов при принятии групповых решений. Предложен метод расчета коэффициента доверия к мнению эксперта при проведении опроса с помощью метода анализа иерархий.

Ключевые слова: экспертная комиссия, принятие групповых решений, метод анализа иерархий, компетентность эксперта.

ВВЕДЕНИЕ

Для повышения степени объективности и качества процедуры принятия решений целесообразно учитывать мнения нескольких экспертов. С этой целью проводится групповая экспертиза, в процессе которой каждый из участников может иметь свое видение решения поставленной задачи или оценку возникшей проблемы [1].

Непосредственно для получения оценок экспертов могут использоваться различные опросы [2]. Одним из способов выявить предпочтения и представить их в количественном виде является метод анализа иерархий (МАИ).

Метод анализа иерархий [3] предполагает декомпозицию проблемы на все более простые составляющие части и обработку суждений лица, принимающего решение. В результате определяется относительная значимость исследуемых альтернатив для всех критериев, находящихся в иерархии.

Относительная значимость выражается численно в виде векторов приоритетов. Полученные таким образом значения векторов являются оценками в шкале отношений и соответствуют так называемым жестким оценкам.

После того как получены оценки от каждого эксперта, их необходимо агрегировать и рассчитать суммарную оценку, которая и будет считаться окончательной.

Поскольку компетентность экспертов в проблеме экспертизы может иметь разную степень, то следует считать мнение более компетентных экспертов более значимым [4]. Для этого используется коэффициент доверия к мнению эксперта.

Коэффициент доверия – это число, которое означает вероятность или степень уверенности, с которой можно считать эксперта компетентным в решаемой проблеме.

1. РАСЧЕТ КОЭФФИЦИЕНТА ДОВЕРИЯ К МНЕНИЮ ЭКСПЕРТА

Сравнение элементов иерархии осуществляется методом попарных сравнений, сущность которого состоит в том, что путем сравнения каждого объекта со всеми другими из данного множества определяются элементы матрицы A размерности $n \times n$, где элемент a_{ij} есть соответствующее действительное число, которое определяет результат сравнения объекта i с объектом j относительно некоторого их общего критерия.

Этап сравнения следует проводить для всех уровней иерархии. В результате будут сформированы матрицы попарных сравнений (МПС) критериев одного уровня иерархии между собой и альтернатив относительно критериев.

Главным недостатком МАИ является то, что процесс заполнения МПС довольно длительный. При наличии m критериев и n альтернатив общее количество сравнений S, которое необходимо выполнить, составляет

$$S = m \cdot (m + n^2). \qquad (1)$$

Для сокращения количества сравнений следует учесть особенности МПС. Все элементы матрицы A положительны: $a_{ij} > 0$ для всех $i, j = 1, \ldots, n$, диагональные элементы a_{ij} должны быть равны единице, так как они выражают оценку критерия относительно самих себя. Поскольку элементы матрицы A являются обратносимметричными $a_{ij} = 1/a_{ij}$ для всех $i, j = 1, \ldots, n$, то эксперт может заполнить только часть матрицы, находящуюся над главной диагональю, а остальная часть значений будет рассчитана математически.

Исходя из этих особенностей, матрица сравнений имеет вид

$$A = \begin{pmatrix} 1 & a_{12} & \dots & a_{1n} \\ 1/a_{12} & 1 & \dots & 1/a_{2n} \\ \dots & \dots & \dots & \dots \\ 1/a_{n1} & 1/a_{n2} & \dots & 1 \end{pmatrix}. \quad (2)$$

Количество сравнений составляет

$$S = \frac{m}{2} \cdot (n^2 - n + m - 1). \quad (3)$$

Для вычисления коэффициента доверия к оценке эксперта предложено добавить в последовательность попарных сравнений серию «контрольных» вопросов. Введение дополнительных вопросов увеличит количество сравнений, которое станет равным

$$S = \frac{m}{2} \cdot (n^2 - n + m + 1). \quad (4)$$

Поскольку при создании МПС часть под главной диагональю рассчитывается как обратная части, заполненной экспертом, то возможно проверить, совпадает ли мнение эксперта при сравнении случайной пары альтернатив с мнением, высказанным им при заполнении матрицы.

Для каждой k-й МПС выбирается случайная пара альтернатив j и i из части матрицы, расположенной под главной диагональю, и предлагается эксперту для сравнения.

Если значения a_{ij} и a_{ji} одновременно либо больше, либо меньше 1, то коэффициент доверия для данного вопроса $v_k = 0$, в противном случае коэффициент доверия для данного вопроса следует рассчитывать по формуле (5).

Пусть $a_{ij} > 1$, а $a_{ji} < 1$, тогда

$$v_k = 1 - \frac{|a_{ij} - 1/a_{ji}|}{\max(a_{ij}, 1/a_{ji})}. \quad (5)$$

Суммарный коэффициент доверия v_s к мнению s-го эксперта вычисляется как

$$v_s = \frac{\sum_{k=1}^{m} v_k}{m}. \quad (6)$$

Если после прохождения опроса значение v_s осталось равным 1, значит, мнению эксперта можно доверять, а если $v_s = 0$, то эксперт некомпетентен в данной проблеме и его мнение не следует учитывать при расчете общей оценки.

Предлагаемая процедура имеет следующий вид:

Шаг 1. Формируется МПС критериев размерностью $m \times m$. Часть над главной диагональю заполняет эксперт, часть под главной диагональю рассчитывается как обратная ей.

Шаг 2. Формируются МПС альтернатив по критериям. Шаги 2.1–2.2 повторяются для каждого критерия.

Шаг 2.1. Заполняется МПС альтернатив по текущему критерию размерностью $n \times n$. Часть над главной диагональю заполняет эксперт, часть под главной диагональю рассчитывается как обратная ей.

Шаг 2.2. Выбирается случайная пара альтернатив x и y из части матрицы, расположенной под главной диагональю, и предлагается эксперту для сравнения. На основе сделанной оценки рассчитывается коэффициент доверия v_k для данного вопроса.

Когда все m критериев рассмотрены, следует перейти к шагу 3.

Шаг 3. Рассчитывается суммарный коэффициент доверия v_s.

Шаг 4. Вычисляются локальные векторы приоритетов W для каждой МПС.

Шаг 5. Определяется вектор глобальных приоритетов GW.

Результатом выполнения этих шагов являются вектор глобальных приоритетов GW и коэффициент доверия к мнению эксперта v_s.

2. ВЫЧИСЛЕНИЕ СУММАРНОЙ ОЦЕНКИ С УЧЕТОМ КОЭФФИЦИЕНТОВ ДОВЕРИЯ К МНЕНИЮ ЭКСПЕРТОВ

На этапе опроса экспертов для каждого эксперта были получены вектор глобальных приоритетов GW и коэффициент доверия к мнению эксперта v_s.

Расчет суммарной оценки Sum каждой i-й альтернативы включает суммирование оценок W_i^s, присвоенных ей каждым из S экспертов. При этом оценки следует умножать на коэффициент доверия к мнению эксперта:

$$Sum_i = \sum_{s=1}^{S} v_s \cdot W_i^s. \quad (7)$$

Таким образом, чем больше коэффициент доверия v_k, тем большее влияние имеет k-й эксперт на общую оценку.

После суммирования вычисляется доля D каждой i-й альтернативы в общей сумме оценок, причем $\sum_{i=1}^{n} D_i = 1$:

$$D_i = \frac{Sum_i}{\sum_{j=1}^{n} Sum_j}. \quad (8)$$

Альтернативы упорядочиваются по убыванию D_i. Лучшей считается альтернатива, чья доля является наибольшей.

3. ВЛИЯНИЕ КОЭФФИЦИЕНТА ДОВЕРИЯ НА СУММАРНУЮ ОЦЕНКУ ТЕНДЕРНОЙ КОМИССИИ

Использование метода продемонстрировано на примере работы экспертной комиссии при проведении тендера.

Тендер – конкурентная форма размещения заказов на выполнение работ по заранее объявленным в документации условиям, в оговоренные сроки на принципах состязательности, справедливости и эффективности. Контракт заключается с победителем тендера – участником, подавшим предложение, соответствующее требованиям документации, в котором предложены наилучшие условия.

Для организации и проведения процедур закупок образуется тендерный комитет на принципах коллегиальности в принятии решений, отсутствия конфликта интересов членов тендерного комитета и их беспристрастности.

В данном примере экспертам было выдано техническое задание к тендеру, включающее пожелания заказчика, и предложения от четырех потенциальных подрядчиков:

– АО «Металлист СМК»;

– Атлас Ворд (Германия);

– IBT (Саудовская Аравия);

– Borga Hale.

Каждый участник комиссии прошел опрос с помощью специального программного обеспечения [5].

Результаты проведения тендера приведены на рис. 1. Каждому эксперту соответствует коэффициент доверия, рассчитанный согласно ответам на контрольные вопросы. На рис. 2 приведены подробные результаты опроса одного из экспертов и сводные матрицы сравнений с глобальными векторами приоритетов.

Как видно из результатов, мнения экспертов относительно победителя не совпадают, и конечный результат определяется оценками экспертов и коэффициентом доверия к мнению каждого эксперта.

Без учета коэффициента доверия рейтинг участников имеет вид:

– АО «Металлист СМК» 32,10 %;

– Borga Hale 25,39 %;

– IBT (Саудовская Аравия) 21,73 %;

– Атлас Ворд (Германия) 20,78 %.

С учетом коэффициентов доверия рейтинг участников имеет вид:

– АО «Металлист СМК» 32,76 %;

– Borga Hale 25,82 %;

– Атлас Ворд (Германия) 21,06 %;

– IBT (Саудовская Аравия) 20,36 %.

Чем сильнее разница в коэффициентах доверия к мнению экспертов, тем большее влияние они будут иметь на конечный результат тендера.

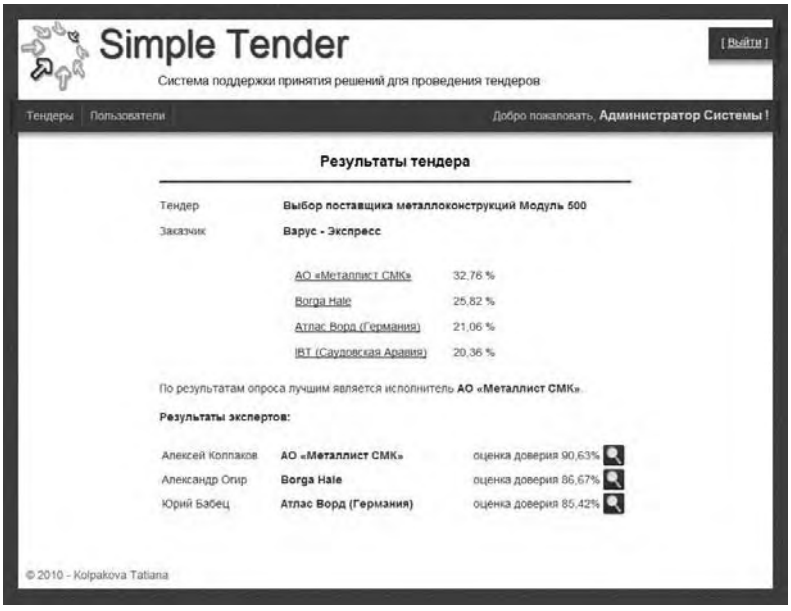
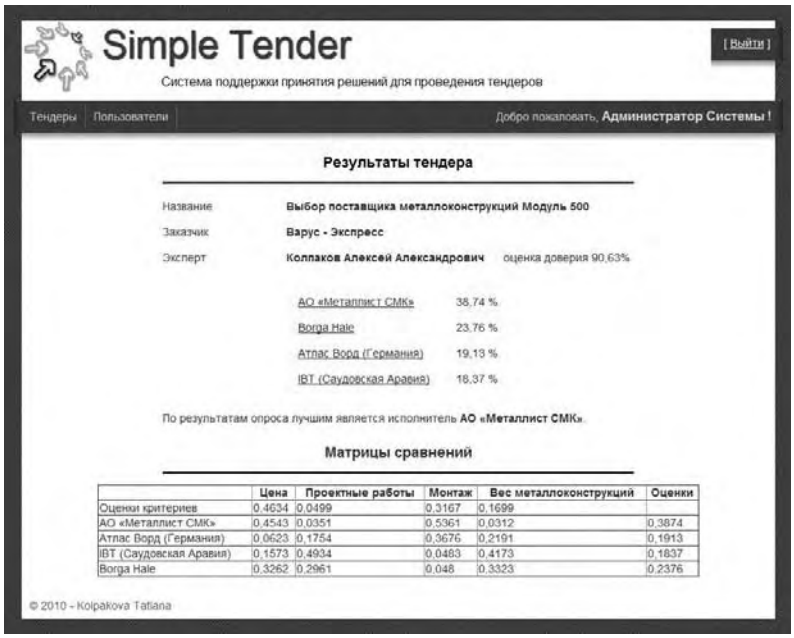


Рис. 1. Результаты проведения тендера



	Цена	Проектные работы	Монтаж	Вес металлоконструкций	Оценки
Оценки критериев	0,4634	0,0499	0,3167	0,1699	
АО «Металлист СМК»	0,4543	0,0351	0,5361	0,0312	0,3874
Атлас Ворд (Германия)	0,0623	0,1754	0,3676	0,2191	0,1913
IBT (Саудовская Аравия)	0,1573	0,4934	0,0483	0,4173	0,1837
Borga Hale	0,3262	0,2961	0,048	0,3323	0,2376

Рис. 2. Результаты опроса эксперта

ВЫВОДЫ

Преимущества разработанного метода:

– в отличие от способов определения «веса» мнения эксперта на основе документационного метода либо методов взаимооценки экспертов [6, 7], данный метод позволяет получить количественное значение коэффициента доверия к мнению эксперта непосредственно в процессе проведения опроса;

– количество сравнений, которые необходимо выполнить для заполнения всех МПС, значительно меньше при использовании расчета части значений. К примеру, для принятия решения по тендеру средней сложности (5 критериев, 4 предложения) необходимо выполнить 105 сравнений, чтобы заполнить все матрицы. С использованием расчета значений МПС количество необходимых сравнений сократится до 40, то есть уменьшится в 2,7 раза.

Недостаток метода заключается в том, что достаточную точность коэффициента доверия к мнению эксперта возможно получить при количестве критериев m не меньше 4, но поскольку в реальных тендерах количество значимых критериев обычно больше четырех, этот недостаток можно считать незначительным.

СПИСОК ЛИТЕРАТУРЫ

1. *Yan, J.* A model of a decision support system based on case-based reasoning for third-party logistics evaluation / Jianyuan Yan, P. E. Chaudhry, Sohail S. Chaudhry // Expert Systems. – 2003. – Vol. 20, Issue 4. – P. 196–207.
2. *Литвак, Б. Г.* Экспертная информация. Методы получения и анализа / Б. Г. Литвак. – М. : Радио и связь, 1982. – 184 с.
3. *Саати, Т.* Принятие решений. Метод анализа иерархий / Т. Саати. – М. : Радио и связь, 1989. – 316 с.
4. *Андрейчиков, А. В.* Анализ, синтез, планирование решений в экономике / А. В. Андрейчиков, О. Н. Андрейчикова. – М. : Финансы и статистика, 2002. – 368 с.
5. *Колпакова, Т. А.* Система поддержки принятия решений для выбора победителя строительного тендера / Т. А. Колпакова // 17-та міжнародна конференція з автоматичного управління «Автоматика-2010» : тези доповідей. Том 2. – Харків : ХНУРЕ, 2010. – С. 170–171.
6. *Кини, Р. Л.* Принятие решений при многих критериях: предпочтения и замещения / Р. Л. Кини, Х. Райфа. – М. : Радио и связь, 1981. – 560 с.
7. *Ramezani, M.* Design and implementation of a fuzzy expert decision support system for vendor selection / Maryam Ramezani, G. A. Montazer // International Conference on Enterprise Information Systems. – 2006. – P. 243–248.

Надійшла 01.11.2010

Колпакова Т. О.

ВИЗНАЧЕННЯ КОМПЕТЕНТНОСТІ ЕКСПЕРТІВ ПРИ ПРИЙНЯТТІ ГРУПОВИХ РІШЕНЬ

Розглядається проблема урахування компетентності експертів при прийнятті групових рішень. Запропоновано метод розрахунку коефіцієнта довіри до думки експерта при проведенні опитування за допомогою методу аналізу ієрархій.

Ключові слова: експертна комісія, прийняття групових рішень, метод аналізу ієрархій, компетентність експерта.

Kolpakova T. A.

DETERMINATION OF EXPERTS COMPETENCE IN GROUP DECISION-MAKING

The problem of taking into account the competence of experts in group decision-making is considered. The method for calculating the coefficient of confidence in the expert's opinion in the survey using the analytic hierarchy is suggested.

Key words: expert committee, group decision-making, analytic hierarchy process, competence of the expert.

УДК 004.8:658

Льовкін В. М.[1], Дубровін В. І.[2], Оніщенко В. Ф.[3]

[1]Аспірант Запорізького національного технічного університету
[2]Канд. техн. наук, професор Запорізького національного технічного університету
[3]Канд. техн. наук, доцент Запорізького національного технічного університету

ПРОГНОЗУВАННЯ ФАКТИЧНИХ РЕЗУЛЬТАТІВ ПРОЕКТУ НА СТАДІЇ ПЕРЕДПРОЕКТНОГО ПЛАНУВАННЯ

Розглянуто проблему прогнозування відхилення фактичних результатів проекту від запланованих на стадії передпроектного планування. Досліджено ефективність застосування моделей на основі нейронних мереж. Запропоновано використання нейро-еволюційного підходу для розв'язання проблеми та значення PDRI і ризику неуспішності проекту як інформативних ознак. Запропоновано критерій інформативності результатів. Представлено метод прогнозування відхилення фактичних результатів проекту від запланованих на стадії передпроектного планування.

Ключові слова: управління проектами, передпроектне планування, прогнозування результатів проекту, Project Definition Rating Index, ризик неуспішності проекту.

ВСТУП

Інвестиційний проект – інвестиційна акція, що передбачає вкладання визначеної кількості ресурсів, у тому числі інтелектуальних, фінансових, матеріальних, людських, для отримання запланованого результату та досягнення встановлених цілей у визначені строки [1]. Для досягнення бажаних результатів у встановлені строки та в межах визначених витрат грошових чи інших важливих ресурсів проекти повинні досконало плануватися та якісно управлятися.

На практиці помилки у відборі проектів, аналізі ризиків та концептуальному плануванні призводять до таких наслідків: обмежені ресурси використовуються на явно неефективні операції; фінансовий, технологічний та конкурентний ризик організації збільшується до неприйнятного рівня [2]. Помилки планування та виконання проектів мають такі наслідки:

– очікуваний прибуток від комерційних контрактів обертається збитками через перевищення початкової вартості, недотримання строків та виплати штрафів;

– затримується введення в дію основних засобів, що призводить до невиконання бізнес-цілей за лінійками продуктів, для яких передбачуються ці засоби;

– проекти за інформаційними системами виконуються з порушенням графіку та перевищенням бюджету, що негативно впливає на управління, загальні витрати та ефективність діяльності, тощо.

Для прийняття рішення інвестор повинен мати за критерії плановані значення основних показників проекту, представлені в документації на проект. Однак фактичні значення в результаті виконання проекту відхиляються від запланованих. Масштабні відхилення можуть призвести до великих втрат або банкрутства. Тому для інвестора або підприємства, яке розглядає декілька варіантів реалізації проекту, дуже корисними для прийняття рішення є дані про прогнозовані значення ризику неуспішності проекту і відхилення розміру витрат та тривалості проекту від запланованих, отримані ще на стадії передпроектного планування, перед початком виконання проекту.

Проблема прогнозування відхилень обсягу витрат та тривалості проекту від запланованих на стадії передпроектного планування є *актуальною* через те, що отримані на даній стадії роботи над проектом значення відхилень, а отже і значення фактичного обсягу та тривалості реалізації, дозволяють ефективно управляти проектом на ранній стадії, не тільки зменшуючи найбільш критичні фактори, які можуть призвести до відхилення, але й ефективно управляючи ресурсами. Управління проектом та ресурсами, зважаючи на отримані прогнозовані значення, дозволяє безпосередньо зменшити відхилення витрат, уникнути нарахування штрафів за невчасне виконання, простоювання чи невчасного забезпечення ресурсами, що також збільшує витрати.

Інструментарій Project Definition Rating Index (PDRI) дозволяє на єдиній основі встановити рівень завершеності визначення масштабу будівельного [3] або промислового проекту [4] перед його виконанням і допомагає передбачати фактори, які викликають ризики проекту. У роботах [5, 6] досліджується

залежність відхилення фактичних витрат та тривалості проекту від PDRI за допомогою регресійної моделі, в [7] запропоновано використання нейронних мереж для прогнозування, результати якого досліджуються порівняно з регресійною моделлю.

У попередніх роботах, де розглядалась дана проблема, авторами було досліджено ефективність застосування різних моделей на основі нейронних мереж та їх ансамблів для прогнозування успішності проектів [8], запропоновано використовувати значення PDRI та запланованих витрат проекту для прогнозування ризику неуспішності проекту, розроблено метод прогнозування ризику неуспішності проектів на стадії передпроектного планування на основі ансамблів нейронних мереж з кластеризацією.

У процесі дослідження в рамках даної роботи необхідно *розв'язати проблему* підвищення ефективності прогнозування фактичних результатів проекту на стадії передпроектного планування за допомогою засобів штучного інтелекту.

1. УПРАВЛІННЯ ПРОЕКТАМИ ТА ПРОГНОЗУВАННЯ ВИТРАТ І ТРИВАЛОСТІ ВИКОНАННЯ ПРОЕКТУ

З метою дослідження існуючих рішень проблеми, що розглядається, був проведений патентний пошук, у результаті якого виявлено декілька патентів, що можуть бути віднесені до даної проблеми.

Патент [9] видано на метод та систему управління проектами, що включають множинність технологічних лістингів, шаблонів робіт та зв'язків серед технологічних лістингів. За допомогою методу та системи забезпечується процес управління проектом, що охоплює побудову моделі управління проектом, яка містить сутності та зв'язки, описані текстовими та графічними даними; введення даних про управління проектом до бази даних відношень; побудову інструментарію управління проектом, що має веб-сторінки з текстовими та графічними даними, генерує гіперпосилання на веб-сторінках, грунтуючись на відносинах у відповідній базі даних. Також у даному патенті представлено інструментарій управління проектом, що охоплює множинність технологічних лістингів, де кожний лістинг забезпечує керівництво щодо того, як виконати діяльність; множинність шаблонів робіт, де кожний шаблон описує реакцію на ситуацію з управління проектом і має зв'язки у лістингах, тощо.

Систему обробки даних та метод використання даної системи для оцінки і управління ризиками запропоновано в патенті [10]. Віддається перевага втіленню методу, що включає кроки з ідентифікації множини елементів ризику, визначаючи важливість кожного виділеного елементу ризику, встановлюючи кожний підризик, що стосується виділених елементів із множини ризиків, визначаючи одну чи більше процедур управління для кожного елементу підризику, встановлюючи ваги для кожної такої процедури. Винахід належить до бізнес-діяльності, а саме до апарату обробки даних та методу ідентифікації, управління та вимірювання ризиків та асоційованих управлінських процедур.

Патент [11] оформлено на винахід – систему управління ризиками проекту для обчислення того, наскільки впливає корегування технологічного плану на весь проект у той час, коли виконується управління проектом. Для визначення такого впливу запропоновано обчислювати розмір впливу корекції чи модифікації у вигляді розподілу щільності ймовірності, а інформація, в якій сформульовано метод корекції окремого процесу залежно від різноманітних факторів відхилення, називається інформацією про правило, в додаток до введення заданих обмежень, що застосовуються умовно. До складу системи управління ризиками проекту входить база даних, у яку записується визначальна інформація, що містить щонайменше одну порцію інформації, яка включає визначення часу можливого початку процесу робіт, можливого завершення та послідовності самих робіт, та база даних розподілу щільності ймовірностей, в яку записується як розподіл щільності ймовірностей щонайменше одна порція інформації, що включає величину коливань для початку процесу робіт, для закінчення робіт та строків будівництва.

У патенті [12] заявлено права на комп'ютерну систему оцінки витрат проекту будівництва, що містить в собі базу даних попередніх будівельних робіт, та засоби оцінки проекту, що включають у тому числі засоби вибору попередніх робіт, які відповідають визначеним користувачем параметрам проекту, та засоби розрахунку оцінок на основі пошуку даних про фактичний час виконання проекту в базі даних для обраних користувачем проектів і обчислення оцінок початкових витрат та тривалості оцінюваного проекту. Даний метод дозволяє вирішити проблему нестачі деталізованих знань з можливістю швидко, просто та точно отримати необхідні дані про тривалість та витрати раніше виконаних робіт, які відповідають параметрам нового проекту.

Система та метод для забезпечення оцінки витрат проекту з використанням комп'ютерної системи, що є частиною мережі, подані в патенті [13]. Винахід належить до засобів підтримки малого бізнесу, а саме до проектно-орієнтованого малого бізнесу, такого

як незалежні підрядники та субпідрядники. Модуль оцінки витрат включає оцінку матеріалів, робіт, накладних витрат тощо. Даний модуль використовує дані, доступні системі, що стосуються архітектурних планів, технічних вимог до матеріалів, вимог до робіт та робітників, для обчислення оцінки витрат, ґрунтуючись на типах та кількості матеріалів і робіт, які мають бути використані у проекті. Окрім того модуль обчислює накладні витрати, такі як фіксовані адміністративні витрати, страхування, обладнання тощо.

Метод оцінки проектних витрат представлено у патенті [14]. Витрати проекту оцінюються автоматично, використовуючи геометричні дані, отримані з плану конструкторських робіт. Протягом виконання проекту або після його завершення дані про фактичні витрати, що стосуються геометричної інформації, отримуються в електронній формі, і ці дані використовуються для оновлення комп'ютеризованого банку даних про витрати.

Винахід у патенті [15] стосується автоматизованого планування будівництва та системи оцінки витрат і програми проекту будівництва. Система оцінки витрат містить базу даних, де зберігається остання інформація про місцеві та регіональні витрати. Програма аналізує файл з вихідними даними, сформульованими користувачем, і перетворює кодові номери елементів на відповідні витрати, дані про які містяться в базі даних.

Патентний пошук дозволив встановити, що наявні розробки не дозволяють вирішити проблему управління проектами на етапі їх вибору з множини запропонованих альтернатив із можливістю не тільки обчислювати заплановані витрати та програму (тривалість) проекту, але й прогнозувати відхилення витрат та тривалості проекту від запланованих значень, виходячи з визначеності проекту на стадії передпроектного планування.

2. ЗАСОБИ ШТУЧНОГО ІНТЕЛЕКТУ ДЛЯ ЗАДАЧ ПРОГНОЗУВАННЯ

Традиційні методи багатовимірної оптимізації є методами локального пошуку, сильно залежать від вибору початкової точки пошуку та накладають додаткові обмеження на властивості цільової функції оптимізації [16].

Властивості нейронних мереж (НМ) у великій мірі залежать від їх архітектури. Якість розв'язання конкретної задачі може бути суттєво покращена за умов використання ансамблів НМ. В ансамблях НМ дані паралельно обробляються декількома НМ, вихідні сигнали яких далі деяким чином комбінуються в об'єднану оцінку, що переважає за якістю результати, отримані за допомогою локальних мереж, що входять до складу ансамблю. На практиці найбільше розповсюдження отримали два підходи до об'єднання мереж в ансамблі: модульний та заснований на зваженому усередненні, і хоча змістовно вони досить відрізняються один від одного, їх об'єднує те, що обидва вони використовують лінійну комбінацію вихідних сигналів своїх членів у тій чи іншій формі [17].

Множина алгоритмів та методів, які використовують для пошуку рішення еволюційні принципи, об'єднується під загальною назвою – еволюційні алгоритми, одним з основних видів яких є генетичний алгоритм.

З погляду штучних систем обробки інформації генетичний пошук є специфічним методом розв'язання задачі оптимізації, при цьому такий ітераційний пошук адаптується до особливостей цільової функції: нові хромосоми, що з'являються в процесі схрещування, тестують все більш широкі області простору пошуку й переважно розташовуються в області оптимуму, а відносно рідкісні мутації перешкоджають виродженню генофонду, що рівносильне рідкісному, але безперервному пошуку оптимуму в решті областей пошукового простору [18].

Репродуктивний план Холланда [19] – канонічна модель генетичного методу. Схема роботи узагальненого генетичного методу представлена на рис. 1 [18].

Як найважливіші характеристики, що визначають властивості конкретного генетичного алгоритму, можна виділити такі:

– спосіб формування початкової популяції $Wj(0)$;

– кількість особин у початковій популяції $Q(0)$, яка повинна бути достатньо великою, щоб покрити всю область можливих рішень;

– частота кросоверу, що визначає кількість хромосом у поточній популяції, що піддаються схрещуванню;

– ймовірність кросоверу для кожної з хромосом поточної популяції;

– частота мутацій, що визначається кількістю хромосом у поточній популяції, які піддаються змінам;

– частота інверсій, що визначається кількістю хромосом у поточній популяції, які піддаються циклічній перестановці генів;

– параметр зміни поколінь $G(k)$, що визначає частину поточної популяції $P(k)$, яка замінюється на кожній ітерації, при цьому $G(k) = 1$ відповідає заміні всієї популяції у кожному поколінні;

– кількість особин у поточній популяції $Q(k)$;

– стратегія селекції [18].

Рис. 1. Схема роботи узагальненого генетичного методу

Спільне застосування штучних НМ та еволюційних алгоритмів, так званий нейро-еволюційний підхід, дає можливість поєднати гнучкість налаштування НМ та адаптивність еволюційних алгоритмів.

3. ЗАСТОСУВАННЯ ЗАСОБІВ ШТУЧНОГО ІНТЕЛЕКТУ ДЛЯ ПРОГНОЗУВАННЯ ФАКТИЧНИХ РЕЗУЛЬТАТІВ ПРОЕКТУ

Для прогнозування фактичного обсягу витрат та тривалості проекту залежно від рівня PDRI були використані моделі на основі НМ зворотного поширення помилки (НМЗП), каскадних НМ прямого поширення сигналу та зворотного поширення помилки (КНМ), радіально-базисних НМ (РБНМ), рекурентних НМ (РНМ), узагальнених регресійних НМ (УРНМ) та НМ Елмана (НМЕ).

Для навчання досліджуваних моделей кожний раз випадковим чином із загальної вибірки в 78 проектів обирались 50 проектів, а інші проекти формували тестову вибірку. Прогнозування обсягу витрат за тривалості проекту відбувалось окремо.

Похибка результатів прогнозування оцінювалась за допомогою відносного відхилення:

$$\text{ПР} = \left| \frac{\text{Прогнозований результат} - \text{Фактичний результат}}{\text{Фактичний результат}} \right|. \quad (1)$$

Результати застосування моделей на основі НМ та регресійних моделей представлені в табл. 1.

У таблиці наведено результати прогнозування для двох прикладів тестових вибірок та середній результат усіх проведених тестувань. НМЗП з 40 нейронами на першому шарі та 10 на прихованому в першому прикладі продемонструвала найкращий результат серед НМЗП, а в другому випадку – найгірший результат,

Таблиця 1. Результати застосування НМ різної архітектури та регресійних моделей

Модель	Тестова вибірка 1	Тестова вибірка 2	Середній результат
Лінійна регресія	62.7020	31.3653	49.0234
Нелінійна регресія	61.6836	31.0025	48.2398
НМЗП [4 2]	74.6228	29.3755	51.2522
НМЗП [10 4]	52.4915	34.8495	43.3218
НМЗП [40 10]	52.1564	60.6491	51.2139
НМЗП [9 3]	56.4677	32.9217	47.0575
НМЗП [8 4]	54.8676	34.5618	47.5376
НМЗП [16 11]	53.5482	45.3177	44.7576
РБНМ	194.8620	32.2059	162.6385
УРНМ	47.3495	30.3659	39.9692
КНМ [4 2]	54.9039	28.3906	44.6617
КНМ [8 4]	51.3613	40.0109	46.8438
КНМ [16 11]	48.5466	43.8835	44.8284
КНМ [40 10]	36.9270	73.9839	50.4928
НМЕ [4 2]	80.5134	53.9108	71.2971
НМЕ [8 4]	79.3931	77.8812	79.1426
НМЕ [16 11]	79.1779	89.8738	83.2219
НМЕ [40 10]	80.8879	99.4334	84.1553
РНМ [4 2]	72.9384	53.8602	60.1055
РНМ [8 4]	60.0677	27.5416	48.5106
РНМ [16 11]	58.5380	29.9436	47.8124
РНМ [40 10]	58.8528	28.1331	47.5710

який перевищив результат найкращої архітектури в даному випадку більш ніж в 2 рази. Така ж тенденція залежності точності результатів прогнозування від навчальної вибірки зберігається для НМ різних архітектур. Зважаючи на дану тенденцію, середній результат застосування моделей на основі лінійної та нелінійної регресії виявився досить ефективним порівняно з моделями на основі НМ.

Результати прогнозування РБНМ та НМЕ виявились досить неточними порівняно з результатами застосування інших підходів.

Однією з задач нейро-еволюційного підходу є еволюційне налаштування структури НМ, при цьому в хромосомі кодується архітектура мережі [20]. В такому випадку для навчання можуть використовуватись градієнтні алгоритми. Пристосованість кожної особини, яка представляє структуру мережі, оцінюється в залежності від результатів навчання. До того ж, нейро-еволюційний підхід в даному випадку дозволяє автоматизувати процес вибору архітектури НМ з можливістю розгляду нерегулярних архітектур, що може привести до отримання кращого результату.

При цьому під час пошуку найефективнішої архітектури для оцінки пристосованості конкретної особини можна проводити навчання мереж відповідної архітектури на основі стандартних методів, а для зменшення впливу випадкових факторів на оцінку структури НМ (адже результати навчання НМ чуттєві до початкових умов та значень параметрів алгоритму навчання) проводити декілька незалежних операцій навчання, а вже найкращий (або усереднений) результат використовувати як оцінку топології НМ – пристосованості особини.

На подальшому етапі, після вибору архітектури НМ для конкретного випадку, з метою розв'язання проблеми обчислення ваг зв'язків НМ вирішено було використати нейро-еволюційний підхід. В даному випадку оптимізуються ваги мереж, значення яких кодуються в хромосомах [20]. Для представлення ваг НМ був обраний дійсний спосіб кодування, тобто хромосома кожної особини була представлена вектором дійсних параметрів.

Однак, у процесі використання нейро-еволюційного підходу для пошуку ваг НМ було виявлено, що в багатьох випадках результати прогнозування НМЗП знаходились майже на одній прямій, а в деяких випадках взагалі вироджувались у пряму $y = b$, де b – константа (рис. 2).

Звичайно, що такі результати прогнозування, які оптимізують середнє відхилення прогнозованих значень від фактично отриманих результатів проекту, є абсолютно неінформативними для осіб, які приймають рішення щодо проекту. Перед авторами постала задача знаходження критерію інформативності, який би дозволив, окрім відхилення значень прогнозованих результатів від фактичних, оцінити ще й інформативність отриманих даних.

Результати проекту можна розглядати у вигляді ламаної лінії (рис. 2), де абсцисі відповідає номер проекту. Відхилення фактичного результату i-го проекту відносно $(i-1)$-го можна обчислити таким чином:

$$B_{\text{ф}} = P^{i}_{\text{ф}} - P^{i-1}_{\text{п}}, \qquad (2)$$

$$B_{\text{п}} = P^{i}_{\text{п}} - P^{i-1}_{\text{п}}, \qquad (3)$$

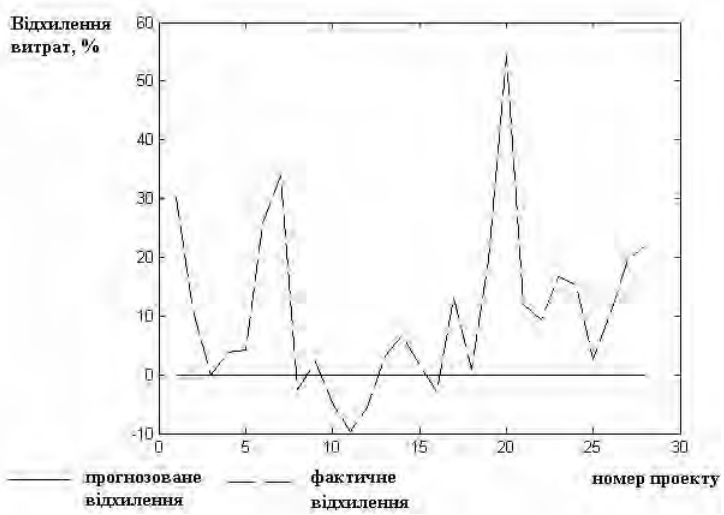


Рис. 2. Приклад прогнозування витрат проекту на основі нейро-еволюційного підходу

де $P_{\text{ф}}^{i}$ – фактичний результат i-го проекту; $P_{\text{п}}^{i}$ – прогнозований результат i-го проекту, $P_{\text{п}}^{i-1}$ – прогнозований результат $(i-1)$-го проекту.

Для розв'язання сформульованої проблеми запропоновано критерій інформативності, який розраховується таким чином:

$$k_{\text{інф}} = \sum_{i=2}^{n} k_{\text{інф}}^{i}, \qquad (4)$$

де

$$k_{\text{інф}}^{i} = \begin{cases} 0, \text{ якщо } \operatorname{sign}(B_{\text{ф}}) \neq \operatorname{sign}(B_{\text{п}}); \\ \dfrac{B_{\text{ф}}}{B_{\text{п}}}, \text{ якщо } \operatorname{sign}(B_{\text{ф}}) = \operatorname{sign}(B_{\text{п}}) \text{ і } B_{\text{ф}} < B_{\text{п}}; \\ \dfrac{B_{\text{п}}}{B_{\text{ф}}}, \text{ якщо } \operatorname{sign}(B_{\text{ф}}) = \operatorname{sign}(B_{\text{п}}) \text{ і } B_{\text{ф}} \geq B_{\text{п}}. \end{cases} \qquad (5)$$

Даний критерій дозволяє оцінити за допомогою тангенса кута нахилу те, наскільки відрізняється прогнозований результат i-го проекту відносно $(i-1)$-го порівняно з тим, наскільки відрізняється фактичний результат i-го проекту відносно прогнозованого результату $(i-1)$-го проекту. При цьому, якщо знак тангенса кута нахилу прогнозованого відхилення не відповідає тангенсу кута нахилу фактичного відхилення, то результат їх співвідношення не враховується, що дозволяє оцінити не оптимізоване значення результату, а його інформативність для особи, що приймає рішення.

4. МЕТОД ПРОГНОЗУВАННЯ ВІДХИЛЕННЯ ФАКТИЧНИХ РЕЗУЛЬТАТІВ ПРОЕКТУ ВІД ЗАПЛАНОВАНИХ НА СТАДІЇ ПЕРЕДПРОЕКТНОГО ПЛАНУВАННЯ

Грунтуючись на результатах дослідження застосування НМ та нейро-еволюційного підходу для прогнозування фактичного обсягу та тривалості проекту, запропоновано метод прогнозування відхилення фактичних результатів проекту від запланованих на стадії передпроектного планування на основі такої процедури:

Крок 1. Сформувати навчальну та тестов і вибірки у вигляді $(\{x_i^1, x_i^2\}, y_i)$, де x_i^1 – значення показника PDRI i-го проекту, x_i^2 – ризик неуспішності i-го проекту (розрахований на основі методу прогнозування ризику неуспішності проектів на стадії передпроектного планування на основі ансамблів НМ з кластеризацією), y_i – відхилення фактичного обсягу витрат (тривалості) i-го проекту від запланованого, %.

Крок 2. Виділити кластери на основі всіх наявних даних за допомогою карт самоорганізації Кохонена відповідно до двох параметрів: показника PDRI та ризику неуспішності проекту, сформувавши K кластерів.

Крок 3. Для кожного кластеру $k = 1\ldots K$ виконати кроки 4–6.

Крок 4. Виділити з навчальної вибірки k-ту навчальну підвибірку, в яку входять проекти, що належать до k-го кластеру.

Крок 5. На основі k-ї навчальної підвибірки за допомогою генетичного алгоритму вибрати оптимальну архітектуру та обчислити ваги зв'язків каскадної НМ.

Крок 6. Використовуючи каскадну НМ, сформовану на кроці 5, для кожного проекту з тестової вибірки, що належить до кластеру k, визначити прогнозований обсяг фактичних витрат (тривалості) проекту.

Крок 7. Зупинення.

Необхідно зазначити, що крок 5 методу може бути реалізований одночасним підбором архітектури та налаштування ваг зв'язків каскадної НМ або в два кроки: обрати архітектуру за допомогою генетичного алгоритму на основі градієнтних методів навчання, а потім для обраної архітектури налаштувати за допомогою генетичного алгоритму ваги зв'язків НМ.

Схематично запропонований метод представлено на рис. 3.

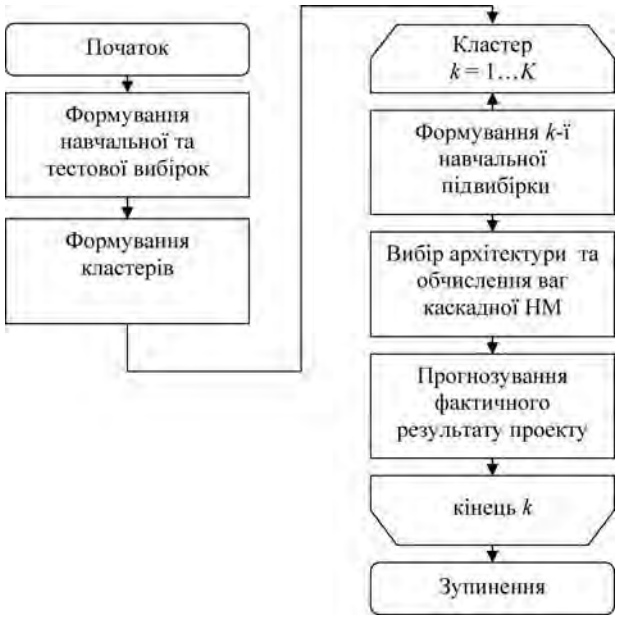


Рис. 3. Метод прогнозування відхилення фактичних результатів проекту від запланованих на стадії передпроектного планування

5. ЕКСПЕРИМЕНТАЛЬНІ РЕЗУЛЬТАТИ ПРОГНОЗУВАННЯ ФАКТИЧНИХ РЕЗУЛЬТАТІВ ПРОЕКТУ НА СТАДІЇ ПЕРЕДПРОЕКТНОГО ПЛАНУВАННЯ

У табл. 2 представлено результати застосування запропонованого методу порівняно з іншими підходами.

Таблиця 2. Результати прогнозування

Підхід	Відносне відхилення	Критерій інформативності
Лінійна регресія	49.0234	6.8006
Нелінійна регресія	48.2398	6.8007
УРНМ	39.9692	7.0352
Нейро-еволюційний підхід на основі НМЗП	25.6915	0.0387
Нейро-еволюційний підхід на основі КНМ	23.065	7.3268
Нейро-еволюційний підхід на основі КНМ з двома інформативними ознаками	21.2372	7.7523
Метод прогнозування відхилення фактичних результатів проекту від запланованих на стадії передпроектного планування	17.6827	8.0205

КНМ продемонстрували кращі результати порівняно з НМЗП за відносним відхиленням (формула (1)), інші моделі на основі НМ (з використанням нейро-еволюційного підходу) виявили гірші результати відносно НМЗП. Запропонований для налаштування архітектури та ваг зв'язків НМ нейро-еволюційний підхід дозволив отримати результат пргнозування, відносне відхилення якого від фактичного перевищує майже в 2 рази відповідний результат на основі звичайних НМ для НМЗП, а для КНМ більш ніж в 2 рази. При цьому результати застосування нейро-еволюційного підходу для НМЗП виявились надто низькими за критерієм інформативності, що вказує на неефективність використання НМЗП для даного підходу. КНМ дозволяють моделювати більш складні зв'язки порівняно з НМЗП через наявність зв'язків не тільки між суміжними шарами. Тому для прогнозування результатів на основі нейро-еволюційного підходу необхідно застосовувати КНМ.

Застосування двох інформативних ознак (значення показника PDRI проекту та його ризику неуспішності) дозволило ще покращити результати прогнозування. Таким чином, запропонований у роботі метод дозволив покращити результати прогнозування на основі автоматизованого налаштування архітектури, ваг зв'язків НМ, використання двох інформативних ознак та кластерів для спеціалізації НМ більш ніж у 2,5 рази порівняно з підходами до розв'язання проблеми, що вже існують [5–7].

ВИСНОВКИ

У роботі розглянуто проблему прогнозування фактичних результатів проекту (витрат і тривалості) на стадії передпроектного планування. Досліджено ефективність застосування моделей на основі нейронних мереж різної архітектури.

Запропоновано використання нейро-еволюційного підходу для вибору архітектури та ваг зв'язків нейронної мережі для прогнозування на основі наявної навчальної вибірки. Це дозволило не тільки автоматизувати даний процес, але й досягти значного покращення середніх результатів моделей на основі НМ, а порівняно з регресійними моделями – у 2 рази. Застосування замість однієї інформативної ознаки двох (значення показника PDRI та ризику неуспішності проекту) і спеціалізації НМ за рахунок кластеризації також дозволило покращити результати.

Запропоновано критерій інформативності, на основі якого оцінено інформативність результатів прогнозування різних підходів.

Наукова новизна роботи полягає в тому, що в роботі запропоновано метод прогнозування відхилення фактичних результатів проекту від запланованих на стадії передпроектного планування, який дозволив підвищити точність результатів прогнозування порівняно з існуючими підходами.

У процесі подальшого дослідження необхідно розробити програмний комплекс для управління інвестиційними проектами на стадії передпроектного планування.

СПИСОК ЛІТЕРАТУРИ

1. *Мазур, И. И.* Управление проектами : учебное пособие / И. И. Мазур, В. Д. Шапиро, Н. Г. Ольдерогге ; под общ. ред. И. И. Мазура. – 5-е изд., перераб. – М. : Омега-Л., 2009. – 960 с.
2. *Арчибальд, Р.* Управление высокотехнологичными программами и проектами / Рассел Д. Арчибальд ; пер. с англ. Мамонтова Е. В. ; под ред. Баженова А. Д., Арефьева А. О. – 3-е изд., перераб. и доп. – М. : Компания АйТи, 2004. – 472 с., ил.
3. *Cho, C.-S.* Building Project Scope Definition Using Project Definition Rating Index / Chung-Suk Cho, G. Edward Gibson Jr. // Journal of Architectural Engineering. – 2001. – Vol. 7, No. 4. – Pp. 115–125.
4. *Gibson, G. E.* Project Definition Rating Index (PDRI) : Construction Industry Institute Research Report / G. E. Gibson, P. R. Dumont. – Austin : UTA, 1996. – 95 p.
5. *Wang, Y.-R.* Applying The PDRI in Project Risk Management : Ph.D. Thesis / Yu-Ren Wang. – Austin, TX, 2002. – 268 p.
6. *Ubach de Fuentes, P.-A.* Validation of the Project Definition Rating Index (PDRI) for MIT Building Projects :

M.S. Thesis / Pere-Andreu Ubach de Fuentes. – Massachusetts Institute of Technology, 2004. – 95 p.
7. *Wang, Y.-R.* A Study of Preproject Planning and Project Success Using ANN and Regression Models / Yu-Ren Wang, G.Edward Gibson Jr. // The 25th International Symposium on Automation and Robotics in Construction. – Vilnius : Vilnius Gediminas Technical University, 2008. – Pp. 688–695.
8. *Дубровін, В. І.* Використання апарату нейронних мереж для прогнозування успішності проектів / В. І. Дубровін, В. М. Льовкін // Сучасні проблеми і досягнення в галузі радіотехніки, телекомунікацій та інформаційних технологій : тези доповідей V Міжнародної науково-практичної конференції (22–24 вересня 2010 р., м. Запоріжжя). – Запоріжжя : ЗНТУ, 2010. – С. 163–165.
9. Pat. 7788118 United States, IPS : G06F 17/50 Project management method and system [Електронний ресурс] / Gerard Vahee, David M. Harris, Jan Heisterberg-Andersen. – Appl. No. : 09/660,852. – Assignee: International Business Machines Corporation ; published : 31.08.2010 ; filed : 13.09.2000. – Режим доступу : http://patimg1.uspto.gov/.piw?Docid=7788118.
10. Pat. 7603283 United States, IPC : G06F 17/50 Method and system for managing risk [Електронний ресурс] / Craig Spielmann, Maria Hutter, Joel Klein, Naresh Singhani. – Appl. No. : 11/783831. – Assignee: JPMorgan Chase Bank, N.A. ; published : 13.10.2009 ; filed : 12.04.2007. – Режим доступу : http://patimg1.uspto.gov/.piw?Docid=7603283.
11. Pat. 7318039 United States, IPC : G06F 9/44 Project risk management system utilizing probability distributions [Електронний ресурс] / Takeshi Yokota, Hisanori Nonaka, Kenji Araki et al. – Appl. No. : 10/246,690. – Assignee: Hitachi Plant Technologies, Ltd. ; published : 8.01.2008 ; filed : 19.09.2002. – Режим доступу : http://patimg1.uspto.gov/.piw?Docid=7318039.
12. Pat. 5918219 United States, IPC : G06Q 10/00 System and method for estimating construction project costs and schedules based on historical data [Електронний ресурс] / John Philip Isherwood. – Appl. No. : 08/357417. – published : 29.06.1999 ; filed : 14.12.1994. – Режим доступу : http://patimg1.uspto.gov/.piw?Docid=5918219.
13. Pat. WO/2001/067335, IPC : G06Q 10/00 System and method of providing project cost evaluation [Електронний ресурс] / Arvin Weiss. – Appl. No. : PCT/US2000/042339. – Applicant: FAIRFAX EXPRESS CORP. ; published : 13.09.2001 ; filed : 29.11.2000. – Режим доступу : http://www.wipo.int/pctdb/en/index.jsp.
14. Pat. WO/2006/034541, IPC : G06F 17/50 Method and system for estimating project costs [Електронний ресурс] / Mark Kefford, Simon Lovegrove, Jason Anderssen. – Appl. No. : PCT/AU2005/001484. – Applicants : EXACTAL PTY LTD (All Except US), Mark Kefford, Simon Lovegrove, Jason Anderssen (US Only) ; published : 06.04.2006 ; filed : 27.09.2005. – Режим доступу : http://www.wipo.int/pctdb/en/index.jsp.
15. Pat. WO/2001/067372, IPC : G06Q 10/00 Computer-implemented automated building design and modeling and project cost estimation and scheduling system [Електронний ресурс] / Robert Bruce Wakelam, Henry C. Beck, Bradley Paul Phillips et al. – Appl. No. : PCT/US2001/001451. – Applicants : BECK TECHNOLOGY (All Except US), Robert Bruce Wakelam, Henry C. Beck et al. (US Only) ; published : 13.09.2001 ; filed : 16.01.2001. – Режим доступу : http://www.wipo.int/pctdb/en/index.jsp.
16. *Дубровін, В. І.* Методи оптимізації та їх застосування в задачах навчання нейронних мереж : навчальний посібник / В. І. Дубровін, С. О. Субботін. – Запоріжжя : ЗНТУ, 2003. – 136 с.
17. *Бодянский, Е. В.* Искусственные нейронные сети: архитектуры, обучение, применения / Е. В. Бодянский, О. Г. Руденко // Харьков : ТЕЛЕТЕХ, 2004. – 369 с. : ил.
18. *Субботін, С. О.* Неітеративні, еволюційні та мультиагентні методи синтезу нечітко логічних і нейромережних моделей : монографія / С. О. Субботін, А. О. Олійник, О. О. Олійник ; під заг. ред. С. О. Субботіна. – Запоріжжя : ЗНТУ, 2009. – 375 с.
19. *Holland, J. H.* Adaptation in natural and artificial systems / J. H. Holland. – Ann Arbor : The University of Michigan Press, 1975. – 97 p.
20. *Цой, Ю. Р.* Эволюционный подход к настройке и обучению искусственных нейронных сетей / Ю. Р. Цой, В. Г. Спицын // Нейроинформатика. – 2006. – Том 1, № 1. – С. 34–61.

Надійшла 04.10.2010

Лёвкин В. Н., Дубровин В. И., Онищенко В. Ф.

ПРОГНОЗИРОВАНИЕ ФАКТИЧЕСКИХ РЕЗУЛЬТАТОВ ПРОЕКТА НА СТАДИИ ПРЕДПРОЕКТНОГО ПЛАНИРОВАНИЯ

Рассмотрена проблема прогнозирования отклонения фактических результатов проекта от запланированных на стадии предпроектного планирования. Исследована эффективность применения моделей на основе нейронных сетей. Предложено использовать нейро-эволюционный подход для решения проблемы и значение PDRI и риска неудачи проекта в качестве информативных признаков. Предложен критерий информативности результатов. Представлен метод прогнозирования отклонения фактических результатов проекта от запланированных на стадии предпроектного планирования.

Ключевые слова: управление проектами, предпроектное планирование, прогнозирование результатов проекта, Project Definition Rating Index, риск неудачи проекта.

Lyovkin V., Dubrovin V., Onyshchenko V.

PREDICTION OF PROJECT ACTUAL RESULTS AT PRE-PROJECT PLANNING STAGE

The problem of predicting project actual results deviation from the results expected at the pre-project planning stage is considered. The efficiency of neural network-based models is analyzed. It is proposed to use the neural-evolution approach for problem solving and to use PDRI and project failure risk values as informative criterions. The criterion of results informativity is proposed. The method of predicting actual project results deviation from the results expected at the pre-project planning stage is presented.

Key words: project management, pre-project planning process, project results prediction, Project Definition Rating Index, project failure risk.

УДК 003.26:004.057.4

Неласа Г. В.[1], Дозоренко І. С.[2]

[1]*Доцент Запорізького національного технічного університету*
[2]*Магістрант Запорізького національного технічного університету*

ОГЛЯД ТА ПОРІВНЯННЯ СХЕМ ЦИФРОВИХ МУЛЬТИПІДПИСІВ

Розглянуто декілька відомих схем цифрових мультипідписів із груповою перевіркою чинності, які використовують лише одну замість кількох перевірок, та недоліки цих схем. Властивості схем порівняно за кількома критеріями.

Ключові слова: цифровий підпис, мультипідпис, групова перевірка, RSA, DSA.

ВСТУП

Цифровий підпис – це назва схожого на традиційний підпис методу, що використовується в криптографії. При використанні традиційного підпису людина пише своє власне ім'я на папері. Ніхто не може підробити інший підпис, тому що важко імітувати чужий почерк. Для впровадження цифрового підпису використовують криптосистеми з відкритим ключем. Кожний підпис має пару ключів: секретний ключ і відкритий ключ. Секретний ключ зберігається у таємниці, у той час як відкритий ключ оприлюднений. Відправник може підписати електронний документ за допомогою цифрового підпису з використанням свого секретного ключа, а одержувач може перевірити цифровий підпис з використанням відкритого ключа відправника. Ніхто не може підробити чужий цифровий підпис, тому що закритий ключ зберігається у таємниці.

Існує багато модифікацій стандартної схеми цифрового підпису, одна з яких, мультипідпис, дозволяє зменшити час верифікації для багатьох підписів. Метою даної роботи є огляд схем цифрових мультипідписів із груповою перевіркою чинності та порівняння їх властивостей.

1. ПОСТАНОВКА ЗАДАЧІ

Традиційно, якщо Аліса хоче відправити повідомлення m, де $m \ p$, Бобу, m повинні бути розділені на t копій $m_1, m_2, \ldots, m_t$. Тоді Аліса підписує ці повідомлення t разів для створення кількох цифрових підписів і відсилає ці повідомлення із цифровими підписами до Боба. Після отримання Боб повинен t разів провести перевірку чинності цих цифрових підписів. Як можна бачити, це потребує багатьох обчислень ступеня за модулем. У випадку використання хеш-функції довжина повідомлення не має значення, але перевірка може займати багато часу через велику кількість повідомлень, що були відправлені.

Для вирішення цієї проблеми Naccache та ін. у 1994 році запропонували схему цифрового мультипідпису із груповою перевіркою [1]. Верифікатор може перевірити цей цифровий мультипідпис за допомогою відкритого ключа, при цьому потрібна тільки одна замість кількох перевірок. Однак, Lim і Lee зазначили [2], що в цій схемі цифровий мультипідпис може бути легко підроблений, для того щоб пройти групову перевірку чинності. У 1998 році Harn запропонував два методи групової перевірки чинності цифрового мультипідпису [3, 4]. Разом з тим, Hwang та ін. зазначили, що ці схеми також не є безпечними [5, 6]. Зловмисник може підробити цифровий мультипідпис для того, щоб пройти групову перевірку чинності. Тому вони запропонували два удосконалення [7]. У 2001 році Shao також запропонував поліпшення для схеми Harn'a [8].

Можна побачити, що якщо цифровий мультипідпис підроблений зловмисником, то має бути перевірений кожний з цифрових підписів. Це означає повернення до вихідної схеми цифрового підпису, яка потребує t перевірок. У 2002 році Changchien і Hwang запропонували алгоритм ефективного виявлення підроблених цифрових мультипідписів [9].

2. СХЕМА NACCACHE ТА ЇЇ НЕДОЛІКИ

Схема мультипідпису Naccache отримана за допомогою модифікації схеми DSA, в якій рівняння перевірки підпису має такий вигляд:

$$r = (g^{ms^{-1}} y^{rs^{-1}} \bmod p) \bmod q, \qquad (1)$$

де $r = g^k \bmod p$, $s = k^{-1}(m + xr) \bmod q$; m – повідомлення; $y = g^x \bmod p$ – відкритий ключ особи, що підписує документ; x – секретний ключ особи, що підписує документ; p – велике просте число; q – великий простий дільник $p - 1$; g – елемент з Z_p порядку q.

Для прискорення перевірки кількох цифрових підписів Naccache та ін. у 1994 році запропонували схему для групової перевірки цифрового мультипідпису. Верифікатор може перевіряти кілька цифрових підписів із використанням відкритого ключа відправника, якому потрібна тільки одна замість t перевірок. Схема виглядає таким чином:

1) Припустимо, що Аліса хоче відправити Бобу t повідомлень $(m_1, m_2, \ldots, m_t)$ та цифровий мультипідпис (r_1, s_1), (r_2, s_2), …, (r_t, s_t). Цифровий мультипідпис був створений за алгоритмом цифрового підпису DSA.

2) Після отримання t повідомлень $(m_1, m_2, \ldots, m_t)$ та мультипідпису (r_1, s_1), (r_2, s_2), …, (r_t, s_t) Боб може перевірити чинність мультипідпису для повідомлень $(m_1, m_2, \ldots, m_t)$ з використанням відкритого ключа Аліси за допомогою рівняння

$$\prod_{i=1}^{t} r_i \bmod p \equiv g^{\sum_{i=1}^{t} -m_i s_i^{-1} \bmod q} \; y^{\sum_{i=1}^{t} r_i s_i^{-1} \bmod q} \bmod p. \quad (2)$$

3) Якщо рівняння виконується, то Боб може стверджувати, що цифровий мультипідпис (r_1, s_1), (r_2, s_2), …, (r_t, s_t) був створений Алісою.

Очевидно, що Боб може зробити групову перевірку цифрового мультипідпису, для чого потрібне тільки одне рівняння (2). Таким чином, схема Naccache та ін. є більш доцільною для групової перевірки кількох цифрових підписів.

Однак Lim і Lee показали, що схема Naccache не є безпечною. Вона має вразливість, завдяки якій зловмисник може підробити цифровий мультипідпис так, щоб рівняння групової перевірки 1.2 виконувалось. Нижче описано спосіб атаки.

1) Зловмисник вибирає довільні числа (u_i, v_i), $i = 1, 2, \ldots, t$ і обчислює $r_i = g^{u_i} y^{v_i} \bmod p$, $i = 1, 2, \ldots, t$.

2) Обчислює $s_b^{-1} \bmod q$, що задовольняє $v_b = r_b s_b^{-1} \bmod q$, $b = 1, 2, \ldots, t$.

3) Зловмисник може отримати s_{t-1} та s_t з рівнянь:

$$\sum_{i=1}^{t} u_i = \sum_{i=1}^{t} m_i s_i^{-1} \bmod q,$$

$$\sum_{i=1}^{t} v_i = \sum_{i=1}^{t} r_i s_i^{-1}.$$

Звідси видно, що зловмисник може підробити t повідомлень $m_1, m_2, \ldots, m_t$ та цифровий мультипідпис (r_1, s_1), (r_2, s_2), …, (r_t, s_t) для того, щоб рівняння групової перевірки (2) виконувалось. Однак жоден з цих підроблених підписів окремо не проходить перевірку в рівнянні (1).

3. СХЕМА ЦИФРОВОГО RSA МУЛЬТИПІДПИСУ HARN'A ТА ЇЇ НЕДОЛІКИ

Схема мультипідпису заснована на алгоритмі RSA, в якій рівняння перевірки підпису має такий вигляд:

$$h(m_i) = S_i^e \bmod n, \quad (3)$$

де $n = p \times q$; $e \times d \bmod (p-1)(q-1) \equiv 1$; p, q – великі прості числа.

Тепер припустимо, що Аліса бажає відправити Бобу t повідомлень $(m_1, m_2, \ldots, m_t)$ та цифровий мультипідпис $S_1, S_2, \ldots, S_t$. Цифровий мультипідпис був створений за алгоритмом RSA, поданим вище. Отже, Аліса відсилає Бобу (m_i, S_i), $i = 1, 2, \ldots, t$.

Після отримання t повідомлень $(m_1, m_2, \ldots, m_t)$ та мультипідпису $S_1, S_2, \ldots, S_t$ Боб може перевірити чинність мультипідпису для повідомлень $(m_1, m_2, \ldots, m_t)$ із використанням відкритого ключа Аліси e за допомогою рівняння

$$\left(\prod_{i=1}^{t} S_i \right)^e = \prod_{i=1}^{t} h(m_i) \bmod n. \quad (4)$$

Якщо рівняння виконується, то Боб може стверджувати, що цифровий мультипідпис $S_1, S_2, \ldots, S_t$ належить Алісі. Очевидно, що Боб може зробити групову перевірку цифрового мультипідпису, для чого потрібне тільки одне рівняння (4). Таким чином, схема Harn'a є більш доцільною для групової перевірки кількох цифрових підписів.

Hwang та ін. запропонували дві атаки для схеми Harn'a. Вони довели, що особа, яка підписує документ, може підробити цифровий мультипідпис так, щоб рівняння групової перевірки (4) виконувалось. Після цього особа може заперечувати, що саме вона підписала ці документи, тобто не виконується умова про неможливість відмови від авторства. Нижче описані способи атаки.

Перший спосіб атаки. Аліса відсилає Бобу підроблені сукупності (m_i, S_i'), $i = 1, 2, \ldots, t$, де $S_i' = h(m_{f(i)})^d \bmod n$, $i = 1, 2, \ldots, t$; $f(\,)$ – бієкція, для якої $f(i) = j$, $i = 1, 2, \ldots, t$ та $j = 1, 2, \ldots, t$. Якщо після отримання підроблених сукупностей (m_i, S_i') Боб перевірить чинність мультипідпису для повідомлень із використанням рівняння групової перевірки (4), то він може стверджувати, що цифровий мультипідпис $S_1', S_2', \ldots, S_t'$ належить Алісі.

Другий спосіб атаки. Аліса відсилає Бобу підроблені сукупності (m_i, S_i'), $i = 1, 2, \ldots, t$, де $S_i' = a_i \times S_i \bmod n$, $i = 1, 2, \ldots, t$ та $\prod_{i=1}^{t} a_i = 1$. Якщо

після отримання підроблених сукупностей (m_i, S_i') Боб перевірить чинність мультипідпису для повідомлень із використанням рівняння групової перевірки (4), то він може стверджувати, що цифровий мультипідпис S_1', S_2', …, S_t' належить Алісі.

Встановлено, що жоден з цих підроблених підписів окремо не проходить RSA перевірку з використанням рівняння (3). Таким чином, Аліса може заперечувати, що вона підписала документи. Схема не відповідає умові про неможливість відмови від авторства.

4. СХЕМА ЦИФРОВОГО DSA-TYPE МУЛЬТИПІДПИСУ HARN'A ТА ЇЇ НЕДОЛІКИ

Схема мультипідпису заснована на алгоритмі DSA-type, що є подібним до алгоритму DSA. Рівняння верифікації має такий вигляд:

$$r = (g^{sr^{-1}} y^{mr^{-1}} \bmod p) \bmod q, \qquad (5)$$

де $r = (g^k \bmod p) \bmod q$; $s = rk - mx \bmod q$; m – повідомлення; $y = g^x \bmod p$ – відкритий ключ особи, що підписує документ; x – секретний ключ особи, що підписує документ; p – велике просте число; q – великий простий дільник $p - 1$; g – елемент з Z_p порядку q.

Для прискорення перевірки кількох цифрових підписів Harn запропонував схему для групової перевірки цифрових підписів. Схема виглядає таким чином:

1) Припустимо, що Аліса хоче відправити Бобу t повідомлень $(m_1, m_2, \ldots, m_t)$ та цифровий мультипідпис (r_1, s_1), (r_2, s_2), …, (r_t, s_t). Цифровий мультипідпис був створений за алгоритмом цифрового підпису DSA-type.

2) Після отримання t повідомлень $(m_1, m_2, \ldots, m_t)$ та мультипідпису (r_1, s_1), (r_2, s_2), …, (r_t, s_t) Боб може перевірити чинність мультипідпису для повідомлень $(m_1, m_2, \ldots, m_t)$ із використанням відкритого ключа Аліси за допомогою рівняння

$$\prod_{i=1}^{t} r_i = \left(g^{\sum_{i=1}^{t} s_i r_i^{-1}} y^{\sum_{i=1}^{t} m_i r_i^{-1}} \bmod p \right) \bmod q. \qquad (6)$$

3) Якщо рівняння виконується, то Боб може стверджувати, що цифровий мультипідпис (r_1, s_1), (r_2, s_2), …, (r_t, s_t) був створений Алісою.

Очевидно, що Боб може зробити групову перевірку цифрового мультипідпису, для чого потрібне тільки одне рівняння (6). Таким чином, схема Harn'a є більш доцільною для групової перевірки кількох цифрових підписів.

Hwang та ін. показали, що схема Harn'a не є безпечною. Вона має вразливість, завдяки якій особа, що підписує документ, може підробити цифровий мультипідпис так, щоб рівняння групової перевірки (6) виконувалось. Після цього особа може заперечувати, що саме вона підписала ці документи, тобто не виконується умова про неможливість відмови від авторства. Нижче описано спосіб атаки.

1) Аліса відсилає Бобу підроблені сукупності (m_i, r_i, s_i'), $i = 1, 2, \ldots, t$, де $s_i' = s_i + a_i r_i \bmod q$, і a_i – ціле число, для якого $\sum_{i=1}^{t} a_i = 0$.

2) Якщо після отримання підроблених сукупностей (m_i, r_i, s_i') Боб перевірить чинність мультипідпису для повідомлень із використанням рівняння групової перевірки (6), то він може стверджувати, що цифровий мультипідпис (r_1, s_i'), (r_2, s_2'), (r_2, s_i') був створений Алісою.

Встановлено, що жоден з цих підроблених підписів окремо не проходить DSA-type перевірку з використанням рівняння (5). Таким чином, Аліса може заперечувати, що вона підписала документи. Схема не відповідає умові про неможливість відмови від авторства, тому що $r \neq (g^{s_i' r_i^{-1}} y^{m_i r_i^{-1}} \bmod p) \bmod q$, $i = 1, 2, \ldots, t$.

5. СХЕМИ ЦИФРОВОГО МУЛЬТИПІДПИСУ HWANG'A

Щоб виправити слабкі сторони схеми цифрового RSA-type мультипідпису Harn'a і схеми цифрового DSA мультипідпису Harn'a, Hwang та ін. запропонували два вдосконалення для цих схем. Перше є поліпшенням схеми цифрового RSA мультипідпису (скорочено BV-RSA). Різниця полягає в рівнянні (3):

$$\left(\prod_{i=1}^{t} S_i^{v_i} \right)^e = \prod_{i=1}^{t} h(m_i^{v_i}) \bmod n, \qquad (7)$$

де v_i, $i = 1, 2, \ldots, t$ є невеликими довільними числами, що обираються перевіряючим.

Друге є поліпшенням схеми цифрового DSA-type мультипідпису (скорочено BV-DSA). Єдина різниця полягає в рівнянні (6). Його модифіковано таким чином:

$$\prod_{i=1}^{t} r_i^{v_i} = \left(g^{\sum_{i=1}^{t} s_i r_i^{-1} v_i} y^{\sum_{i=1}^{t} m_i r_i^{-1} v_i} \bmod p \right) \bmod q, \qquad (8)$$

де v_i, $i = 1, 2, \ldots, t$, є невеликими довільними числами, що обираються перевіряючим.

6. СХЕМА ЦИФРОВОГО DSA-TYPE МУЛЬТИПІДПИСУ SHAO

У 2001 році Shao запропонував схему цифрового DSA-type мультипідпису. Вона схожа на схему Hwang'a та ін. Єдина різниця полягає в рівнянні 6. Рівняння групової перевірки виглядає таким чином:

$$\prod_{i=1}^{t}(e_i(s_i))^{u_i} = \prod_{i=1}^{t}(f_i(s_i))^{u_i} \bmod p, \qquad (9)$$

де $u_i \in (1, 2^{32})$, $i = 1, 2, \ldots, t$ є довільними числами, що обираються перевіряючим, а s_i – цифровий мультипідпис, кожен підпис якого окремо задовольняє рівнянню $e_i(s_i) = f_i(s_i) \bmod p$, $i = 1, 2, \ldots, t$.

7. СХЕМА CHANGCHIEN'A ТА ІНШІ

В розділі 2 зазначено, що якщо не виконується рівняння групової перевірки, тобто $\left(\prod_{i=1}^{t} S_i\right)^e \neq \prod_{i=1}^{t} h(m_i) \bmod n$, отримувач, Боб, має перевірити кожен підпис із мультипідпису окремо з використанням рівняння $h(m_i) = S_i^e \bmod n$. Визначення підробленого підпису потребує t обчислень ступеня. У 2002 році Changchien і Hwang запропонували схему для визначення підроблених мультипідписів, що потребує лише одного обчислення ступеня та t обчислень модуля.

Changchien і Hwang перевизначили $h(\,)$ як просту необоротну хеш-функцію та $\prod_{i=1}^{t} h(m_i) \le n$. Це робить довжину $h(\,)$ рівною $\lfloor |n|/t \rfloor$ біт, де $\lfloor \cdot \rfloor$ – функція найбільшого цілого, а $|n|$ – довжина n. Для того, щоб визначити підроблений мультипідпис, Боб має виконати такі кроки:

1) Обчислити $L = \left(\prod_{i=1}^{t} S_i\right)^e \bmod n$.

2) Перевірити, чи $L \bmod h(m_i) = 0$ для $i = 1, 2, \ldots, t$.

8. ПОРІВНЯННЯ

Порівняти наведені схеми мультипідписів можна за такими критеріями:

1) *Лише чинна особа може підписати електронний документ цифровим мультипідписом.*

Всі схеми відповідають цій умові. Будь-яка особа, що має свій секретний ключ, може це зробити.

2) *Ніхто не може підробити чужий цифровий мультипідпис.*

Серед розглянутих схем лише схеми BV-DSA та BV-RSA Hwang'a, а також схема Shao відповідають цій умові. Мультипідпис за цими схемами неможливо підробити для того, щоб пройти групову перевірку чинності.

3) *Будь-який перевіряючий може провести групову перевірку чинності цифрового мультипідпису.*

Всі схеми відповідають цій умові. Будь-який перевіряючий може провести групову перевірку цього мультипідпису за допомогою відкритого ключа, для цього потрібна тільки одна перевірка.

4) *Контроль цілісності.*

Всі схеми відповідають цій умові. Зловмисник не має змоги замінити дійсний документ фальшивим, бо він не знає секретного ключа особи, що підписала документ. Лише чинна особа може підписувати свої документи.

5) *Неможливість відмови від авторства.*

Якщо відправник може підробити цифровий підпис, що проходить групову перевірку чинності, то схема не відповідає умові, бо відправник може заперечувати, що саме він відправив ці документи. Отже, лише схеми BV-DSA та BV-RSA Hwang'a, а також схема Shao відповідають цій умові.

6) *Має бути ефективний метод виявлення фальшивих цифрових мультипідписів.*

Практично всі схеми, крім схеми Changchien'a та ін., не відповідають цій умові. В цій схемі перевіряючий може ефективно визначити факт підроблення підпису.

Результати порівняння підсумовано у табл. 1, де K_i – визначені критерії, «+» – вирішено, «–» – не вирішено, ТОЗ – тип обчислювальної задачі, ЗДЛ – задача дискретного логарифмування, ЗФ – задача факторизації.

Таблиця 1. Результати порівняння схем мультипідписів

Назва схеми	K_1	K_2	K_3	K_4	K_5	K_6	ТОЗ
Naccache та ін.	+	–	+	+	–	–	ЗДЛ
DSA Harn'a	+	–	+	+	–	–	ЗДЛ
RSA Harn'a	+	–	+	+	–	–	ЗФ
BV-DSA Hwang'a	+	+	+	+	+	–	ЗДЛ
BV-RSA Hwang'a	+	+	+	+	+	–	ЗФ
Shao	+	+	+	+	+	–	ЗДЛ
Changchien'a та ін.	+	–	+	+	–	+	ЗФ

ВИСНОВКИ

Було розглянуто декілька схем цифрових мультипідписів, що вже існують. Ці схеми дозволяють будь-якому перевіряючому проводити групову перевірку чинності цифрових підписів. Вони дозволяють заощадити багато обчислень ступеня за модулем. Тим не менш, лише схеми BV-DSA та BV-RSA Hwang'a, а також схема Shao є надійними та забезпечують умову неможливості відмови від авторства. Однак ці схеми не мають ефективного методу виявлення підроблених підписів на відміну від схеми Changchien'a. Проблема створення безпечної та ефективної схеми цифрового мультипідпису залишається відкритою та може розглядатися як напрямок подальших досліджень.

СПИСОК ЛІТЕРАТУРИ

1. *Naccache, D.* Can DSA be improved: Complexity trade-os with the digital signature standard / D. Naccache, D. Mraihi, D. Rapheali, S. Vaudenay // Proceedings of Eurocrypt'94. – 1994. – Pp. 85–94.
2. *Lim, C. H.* Security of interactive DSA batch verication / C. H. Lim, P. J. Lee // Electronics Letters. – 1994. – Vol. 30, No. 19. – Pp. 1592–1593.
3. *Harn, L.* Batch verifying multiple DSA-type digital signatures // Electronics Letters. – 1998. – Vol. 34, No. 9. – Pp. 870–871.
4. *Harn, L.* Batch verifying multiple RSA digital signatures // Electronics Letters. – 1998. – Vol. 34, No. 12. – Pp. 1219–1220.
5. *Hwang, M. S.* Cryptanalysis of the batch verifying multiple RSA digital signatures / M. S. Hwang, I. C. Lin, K. F. Hwang // Informatica. – 2000. – Vol. 11, No. 1. – Pp. 15–19.
6. *Hwang, M. S.* Cryptanalysis of the batch verifying multiple DSA-type digital signatures / M. S. Hwang, C. C. Lee, Eric J. L. Lu // Pakistan Journal of Applied Sciences. – 2001. – Vol. 1, No. 3. – Pp. 287–288.
7. *Hwang, M. S.* Two simple batch verifying multiple digital signatures / M. S. Hwang, C. C. Lee, and Y. L. Tang // The Third International Conference on Information and Communication Security (ICICS2001). – Xian, China, 2001. – Pp. 13–16.
8. *Shao, Z.* Batch verifying multiple DSA-type digital signatures // Computer Networks. – 2001. – Vol. 37, No. 3–4. – Pp. 383–389.
9. *Changchien, S. W.* A batch verifying and detecting multiple RSA digital signatures / S. W. Changchien, M. S. Hwang // International Journal of Computational and Numerical Analysis and Applications. – 2002. – Vol. 2, No. 3. – Pp. 303–307.

Надійшла 29.10.2010

Неласая А. В., Дозоренко И. С.

ОБЗОР И СРАВНЕНИЕ СХЕМ ЦИФРОВЫХ МУЛЬТИПОДПИСЕЙ

Рассмотрены известные схемы цифровых мультиподписей с групповой проверкой, использующие только одну вместо нескольких проверок, а также недостатки этих схем. Проведено сравнение свойств схем по нескольким критериям.

Ключевые слова: цифровая подпись, мультиподпись, групповая проверка, RSA, DSA.

Nelasa A. V., Dozorenko I. S.

REVIEW AND COMPARISON OF MULTIPLE DIGITAL SIGNATURES

Several batch verication multiple digital signatures are reviewed in this paper. These schemes use only one verification instead of several verifications. Weakness of these schemes is also pointed out. The schemes were compared by the defined criterions.

Key words: multiple digital signatures, batch verification, RSA, DSA.

УДК 681.142.2; 622.02.658.284; 621.325

Пелешко Д. Д.[1], Кустра Н. О.[2], Шпак З. Я.[1]

[1]Канд. техн. наук, доцент Національного університету «Львівська політехніка»

[2]Канд. техн. наук, старший викладач Національного університету «Львівська політехніка»

СУМІЩЕННЯ ЗОБРАЖЕНЬ НАБОРУ НА ОСНОВІ ВИКОРИСТАННЯ ДИСПЕРСІЇ КОЛЬОРУ ЗОБРАЖЕНЬ

Розроблено швидкий метод суміщення зображень в наборі однотипних зображень на основі розв'язання задачі майже факторизації простору топології зображень з подальшим звуженням цього простору через вирішення задачі пошуку кореляційного максимуму. Задача майже факторизації формулюється через введення напівметрики стосовно дисперсії кольору елементів топології зображення.

Ключові слова: суміщення зображень, фреймове покриття, топологія зображень, дисперсія кольору, кореляційний максимум.

ВСТУП

Традиційно для реалізації процедури знаходження і суміщення зображень використовують кореляційну прив'язку цифрових зображень. Метод кореляційної прив'язки зображень має такі недоліки:

– взаємна кореляційна функція може мати досить розмитий максимум, що ускладнює його знаходження, оскільки не враховує просторову структуру порівнюваних зображень;

– комбінаторна складність – великий перебір ситуацій [1–4].

Основу запропонованого методу складають:

– запропоновані в [4] топологічні представлення та операції, зокрема звуження простору покриття зображення.

– характеристики виділених в [3] класів представлення зображень та наборів.

1. ПОСТАНОВКА ЗАДАЧІ

Метою даної роботи є розробка швидкого методу суміщення зображень в наборі на основі використання дисперсії значень кольору (чи інтенсивності).

Для досягнення цієї мети до розгляду потрібно ввести топологію зображення і визначити на ній задачу майже факторизації топологічного простору.

Основна ідея пропонованого методу суміщення полягає у швидкому формуванні для кожного зображення відповідних наборів «підозрілих» на подібність фреймів (задача звуження простору топологічного покриття зображення через вирішення задачі майже факторизації) з подальшим їх звуженням математичною кореляцією із заданим фрагментом (задача звуження простору топологічного покриття зображення через вирішення задачі пошуку кореляційного максимуму на топологічному покритті зображення).

2. ТОПОЛОГІЇ ДЛЯ ЗАДАЧІ СУМІЩЕННЯ ЗОБРАЖЕНЬ НАБОРУ

Нехай задано набір $\mathbf{P}$ однотипних рисунків з координатною $\Im_{\mathbf{P}} = \Im_{\mathbf{X}^{2,+,d}}$ та колірною топологіями $\mho_{\mathbf{P}}$ [5]. При цьому треба пам'ятати, що $\mho_{\mathbf{P}}$ індукується $\Im_{\mathbf{P}}$. В кожній з цих топологій визначимо скінченні покриття: фреймове $(\chi_{\mathbf{X}^{2,+,d}} | N_\chi) \subseteq \Im_{\mathbf{X}^{2,+,d}}$ та індуковане фрагментне $\vartheta_{\mathbf{P}}$ в $\mho_{\mathbf{P}}$ [6].

Серед зображень набору виберемо довільне зображення, стосовно якого буде здійснюватись операція суміщення. Таке зображення будемо називати *фіксованим*. Для зручності подальшого викладу нехай таке зображення має індекс в наборі, рівний 1. Тобто в наборі $\mathbf{P}$ фіксованим є зображення $P_{\text{фікс}} = P_1$. Тоді через $\mathbf{P}'$ позначимо набір з решти зображень

$$\mathbf{P}' = \mathbf{P} \backslash \{P_1\} = \{P_z\}_{z=2\ldots N}. \quad (1)$$

На $(\chi_{\mathbf{X}^{2,+,d}} | N_\chi)$ визначимо фрейм

$$\mathbf{X}^{2,+,d}_{\mathbf{fr}1,\,\text{зад}} = \mathbf{X}^{2,+,d}_{\mathbf{fr}1,\,\text{зад}}(\Delta_{x\,1,\text{зад}}, \Delta_{y\,1,\text{зад}}, l_{\mathbf{fr}1,\,\text{зад}}, h_{\mathbf{fr}1,\,\text{зад}}), \quad (2)$$

якому на P_1 відповідає індукований фрагмент зображення $P_{1,\,\text{зад}} \in \vartheta_{\mathbf{P}}$.

Проблема вибору початкового фрейму $\mathbf{X}^{2,+,d}_{\mathbf{fr}1,\,\text{зад}}$ в даній роботі детально не розглядатиметься. Це питання детально розглядалось в [7]. Приймаємо лише одне припущення – $\mathbf{X}^{2,+,d}_{\mathbf{fr}1,\,\text{зад}}$ індукує такий фрагмент зображення P_1, який з достатньою точністю існує на усіх зображеннях набору $\mathbf{P}'$.

Вважатимемо, що фреймове покриття $(\chi_{\mathbf{X}^{2,+,d}} | N_\chi)$ гомеоморфне фрейму $\mathbf{X}^{2,+,d}_{\mathbf{fr}1,\,\text{зад}}$ за розмірами. Тут гомеорфізм за розмірами визначає те, що усі елементи $(\chi_{\mathbf{X}^{2,+,d}} | N_\chi)$ мають розміри $l_{\mathbf{fr}1,\,\text{зад}}$ і $h_{\mathbf{fr}1,\,\text{зад}}$, а відрізняються лише координатами початку.

З $(\chi_{\mathbf{X}^{2,+,d}} | N_\chi)$ при заданій $\Im_{\mathbf{P}}$ сформуємо фреймове покриття набору $\mathbf{P}'$ за правилом

$$\chi_{\mathbf{P}'} = \{\chi_z\}_{z=2\ldots N},\ (\chi_z | N_\chi) \in \Im_{\mathbf{P}}, \quad (3)$$

де

$$\forall z, z \in [2\ldots N]:\ \chi_{z_1} = \chi_{z_2};\ \chi_{z_1}, \chi_{z_1} \in \chi'_{\mathbf{P}}. \quad (4)$$

Формули (3) і (4) означають, що фреймове покриття $\chi_{\mathbf{P}'}$ набору $\mathbf{P}'$ складається з $N-1$ топологічно еквівалентних покриттів $(\mathbf{X}^{2,+,d}, \Im_{\mathbf{X}^{2,+,d}})$, елементи яких ще й рівні за розмірами. При цьому важливо відзначити, що розмірність кожного χ_z рівна N_χ. Тоді має місце

$$\dim \chi_{\mathbf{P}'} = (N-1)N_\chi, \quad (5)$$

і до розгляду треба приймати топологічний простір $(\chi_{\mathbf{P}'} | (N-1)N_\chi)$.

Фреймове покриття (3) засобом неперервного відображення $\mathbf{C}$ [6] індукує фрагментне покриття $\vartheta_{\mathbf{P}'} \subseteq \vartheta_{\mathbf{P}}$, яке належить топології $\mho_{\mathbf{P}'} \subseteq \mho_{\mathbf{P}}$ набору $\mathbf{P}'$, за правилом

$$\begin{aligned} \mho_{\mathbf{P}'} &= \mho_{\mathbf{P}} \backslash \{\mho_1\} = \{\mho_z\}_{z=2\ldots N}; \\ \vartheta_{\mathbf{P}'} &= \vartheta_{\mathbf{P}} \backslash \{\vartheta_1\} = \{\vartheta_z\}_{z=2\ldots N}; \\ \vartheta_{\mathbf{P}'} &\subseteq \mho'_{\mathbf{P}} \subseteq \mho_{\mathbf{P}} \end{aligned} \quad (6)$$

Фактично $\chi_{\mathbf{P}'}$ і $\vartheta_{\mathbf{P}'}$ виступають звуженнями $\chi_{\mathbf{P}}$ і $\vartheta_{\mathbf{P}}$ відповідно.

Оскільки $\vartheta_{\mathbf{P}}$ є індуковане неперервним відображення $\mathbf{C}$ [6], то визначений для $(\chi_{\mathbf{X}^{2,+,d}} | N_\chi)$ гомеорфізм (за розмірами) до фрейму $\mathbf{X}^{2,+,d}_{\mathbf{fr}1,\,\text{зад}}$ має місце для елементів просторів $\vartheta_{\mathbf{P}}$ і $\vartheta_{\mathbf{P}'}$ до фрагмента $P_{1,\,\text{зад}}$. При цьому для елементів $\vartheta_{\mathbf{P}}$ і $\vartheta_{\mathbf{P}'}$ не існує топологічної еквівалентності, подібної до (3). Це означає, що набір $\mathbf{P}'$ можна подати у вигляді скінченного набору фрагментів $P_{z,m}$ з розмірами $l_{\mathbf{fr}1,\,\text{зад}}$ і $h_{\mathbf{fr}1,\,\text{зад}}$

$$\begin{aligned} \mathbf{P}' &= \{P_z\} = \\ &= \left\{ \left\{ P_{z,m} \middle| P_{z,m} = C_{z,m}(\mathbf{X}^{2,+,d}_{\mathbf{fr}z,m}) \right\}_{m=1\ldots N_\chi} \right\}_{z=2\ldots N}; \\ &\forall z, m : \mathbf{X}^{2,+,d}_{\mathbf{fr}z,m} \in \chi_z. \end{aligned} \quad (7)$$

Зазначимо, що розмірність кожного ϑ_z складає N_χ, тобто існує простір $(\vartheta_z|N_\chi)$. Тоді розмірність $\vartheta_{\mathbf{P}'}$ за (5) складає

$$\dim\vartheta_{\mathbf{P}'} = (N-1)N_\chi, \tag{8}$$

і до розгляду треба приймати простір $(\vartheta_{\mathbf{P}'}|(N-1)N_\chi)$.

3. СУМІЩЕННЯ ЗОБРАЖЕНЬ НАБОРУ НА ОСНОВІ ВИКОРИСТАННЯ ДИСПЕРСІЇ

3.1. Майже факторизація просторів покриття зображень набору на основі дисперсії

Нехай задано набір $\mathbf{P}$, фіксоване зображення $P_{\text{фікс}} = P_1$, фрейм $\mathbf{X}^{2,+,d}_{\mathbf{fr}1,\text{зад}}$ і фрагмент $P_{1,\text{зад}}$, набір $\mathbf{P}'$ (1), топологій $\Im_{\mathbf{P}} = \Im_{\mathbf{X}^{2,+,d}}$ та $\mathfrak{V}_{\mathbf{P}'} \subseteq \mathfrak{V}_{\mathbf{P}}$ [4] і покриття $(\chi_{\mathbf{P}'}|(N-1)N_\chi)$ (3) та $(\vartheta_{\mathbf{P}'}|(N-1)N_\chi)$ (6).

Для кожного фрагмента $P_{z,m}$ визначимо дисперсію $D_{z,m}$ [1] значення кольору (чи інтенсивності) $c^d_{z,m,(i,j)}$ кожного фрагмента. Розрахункова формула має вигляд

$$D_{z,m} = \frac{1}{s_{1,\text{зад}}} \sum_{i=x_{\text{поч}\,z,m}}^{x_{\text{поч}\,z,m}+l_{\mathbf{fr}1,\text{зад}}} \sum_{j=y_{\text{поч}\,z,m}}^{y_{\text{поч}\,z,m}+n_{\mathbf{fr}1,\text{зад}}} \left(c^d_{z,m,(i,j)} - M_{z,m}\right)^2;$$
$$m = 1\ldots N_\chi;$$
$$z = 2\ldots N, \tag{9}$$

де $s_{1,\text{зад}} = l_{\mathbf{fr}1,\text{зад}} h_{\mathbf{fr}1,\text{зад}}$ – площа $P_{1,\text{зад}}$: $M_{z,m}$ – математичне сподівання.

Подібно до (9) обчислюється дисперсія $D_{1,\text{зад}}$ для фрагмента $P_{1,\text{зад}}$.

В результаті (9) кожному фрагменту $P_{z,m}$ однозначно поставлена у відповідність характеристика – середнє значення кольорів $D_{z,m}$ відповідного фрагмента зображення P_z

$$P_{z,m} \to D_{z,m}. \tag{10}$$

Це означає, що $(\chi_{\mathbf{P}'}|(N-1)N_\chi)$ засобами (10) через $(\vartheta_{\mathbf{P}'}|(N-1)N_\chi)$ індукує набір характеристик – дисперсій кольору

$$\chi_{\mathbf{P}'} \underset{\mathbf{C}}{\to} \vartheta_{\mathbf{P}'} \underset{M}{\to} \{D_{z,m}\},$$
$$z = 2\ldots N. \tag{11}$$

Для задачі майже факторизації $(\chi_{\mathbf{P}'}|(N-1)N_\chi)$ введемо напівметрику $d_{D,\mathbf{fr}}(P_{z,m}, P_{1,\text{зад}})$ як відношення еквівалентності фрагменту $P_{1,\text{зад}}$

$$\forall P_{z,m} \in \vartheta_{\mathbf{P}'}: d_{D,\mathbf{fr}}(P_{z,m}, P_{1,\text{зад}}) =$$
$$= |D_{z,m} - D_{1,\text{зад}}|, \tag{12}$$

Твердження. (12) є напівметрикою.

Доведення.

◁

Оскільки (12) є евклідовою відстанню, то звідси випливають умови метрики.

Відношення еквівалентності як умова напівметрики випливає з того, що для дисперсії (9) як інтегральної характеристики фрагмента можлива ситуація, коли

$$\exists z \in [2\ldots N],\ m \in [2\ldots N_\chi]: D_{z,m} = D_{1,\text{зад}}. \tag{13}$$

Це означає, що для $P_{z,m} \neq P_{1,\text{зад}}$ має місце

$$d_{D,\mathbf{fr}}(P_{z,m}, P_{1,\text{зад}}) = 0, \tag{14}$$

що визначає метрику (12) як напівметрику.

▷

Тоді задача майже факторизації простору $(\vartheta_{\mathbf{P}'}|(N-1)N_\chi)$ полягає у побудові $\vartheta_{\left(\vartheta_{\mathbf{P}'}|(N-1)N_\chi\right), d_{D,\mathbf{fr}}} /^{\sim\varepsilon}$ за допомогою нерівності

$$\forall P_{z,m} \in \vartheta_{\left(\vartheta_{\mathbf{P}'}|(N-1)N_\chi\right), d_{D,\mathbf{fr}}} /^{\sim\varepsilon},$$
$$d_{D,\mathbf{fr}}(P_{z,m}, P_{1,\text{зад}}) \le \varepsilon, \tag{15}$$

де ε – точність суміщення – параметр майже факторизації.

В загальному випадку треба розглядати $\varepsilon = \varepsilon(z)$. Проте на практиці для зручності вибирають точність одну для усіх $N-2$ рисунків набору $\mathbf{P}'$.

Фактично $\vartheta_{\left(\vartheta_{\mathbf{P}'}|(N-1)N_\chi\right), d_{D,\mathbf{fr}}} /^{\sim\varepsilon}$ треба розглядати як набір фрагментів набору $\mathbf{P}'$ «підозрілих» на подібність (за 15) фрагменту $P_{1,\text{зад}}$. Оскільки

$$\forall \vartheta_z \in \vartheta_{\left(\vartheta_{\mathbf{P}'}|(N-1)N_\chi\right), d_{D,\mathbf{fr}}} /^{\sim\varepsilon}: \dim\vartheta_z \le N_\chi, \tag{16}$$

то має місце оцінка

$$\dim\vartheta_{\left(\vartheta_{\mathbf{P}'}|(N-1)N_\chi\right), d_{D,\mathbf{fr}}} /^{\sim\varepsilon} \le$$
$$\le \dim\vartheta_{\mathbf{P}'} = (N-1)N_\chi. \tag{17}$$

Якщо прийняти, що $\vartheta_{\left(\vartheta_{\mathbf{P}'}|(N-1)N_\chi\right), d_{D,\mathbf{fr}}} /^{\sim\varepsilon}$ відповідає майже фактор $\chi_{\left(\chi_{\mathbf{P}'}|(N-1)N_\chi\right), d_{D,\mathbf{fr}}} /^{\sim\varepsilon}$, такий що

$$\left(\chi_{\left(\chi_{\mathbf{P}'}|(N-1)N_\chi\right), d_{D,\mathbf{fr}}} /^{\sim\varepsilon} \middle| \vartheta_{\left(\vartheta_{\mathbf{P}'}|(N-1)N_\chi\right), d_{D,\mathbf{fr}}} /^{\sim\varepsilon}\right) \subseteq$$
$$\subseteq (\chi_{\mathbf{P}'}|(N-1)N_\chi), \tag{18}$$

то (17) означає, що через вирішення задачі майже факторизації вдалось звузити простори $(\vartheta_{\mathbf{P}'}|(N-1)N_\chi)$ і $(\chi_{\mathbf{P}'}|(N-1)N_\chi)$ відповідно.

3.2. Задача пошуку кореляційного максимуму на майже фактор просторі зображення

Наступним кроком є звуження просторів $\vartheta_{\left(\vartheta_{\mathbf{P}'}|(N-1)N_\chi\right),\, d_{D,\,\mathbf{fr}}}^{\sim\varepsilon/}$ і $\chi_{\left(\chi_{\mathbf{P}'}|(N-1)N_\chi\right),\, d_{D,\,\mathbf{fr}}}^{\sim\varepsilon/}$ до одного фрейму через вирішення задачі пошуку кореляційного максимуму. Для цього введемо до розгляду метрику

$$\forall P_{z,\,m} \in \vartheta_{\mathbf{P}'} : d_{r_{\max},\,\mathbf{fr}}(P_{z,\,m}, P_{1,\,\text{зад}}) = \\ = r(P_{z,\,m}, P_{1,\,\text{зад}}), \qquad (19)$$

де $r(P_{z,\,m}, P_{1,\,\text{зад}})$ – кореляції [1] між значеннями кольору (чи інтенсивності) фрагменту $P_{z,\,m}$ із заданим $P_{1,\,\text{зад}}$. Розрахункова формула має вигляд

$$r_{z,\,m}(P_{1,\,\text{зад}}, P_{z,\,m}) = \frac{\sum\limits_{i\,=\,x_{\text{поч}\,z,\,m}}^{x_{\text{поч}\,z,\,m}+l_{\mathbf{fr}1,\,\text{зад}}} \sum\limits_{j\,=\,y_{\text{поч}\,z,\,m}}^{y_{\text{поч}\,z,\,m}+h_{\mathbf{fr}1,\,\text{зад}}} (c^d_{1,\,\text{зад},\,(i,\,j)} - M_{1,\,\text{зад}})(c^d_{z,\,m,\,(i,\,j)} - M_{z,\,m})}{\sqrt{\sum\limits_{i\,=\,x_{\text{поч}\,z,\,m}}^{x_{\text{поч}\,z,\,m}+l_{\mathbf{fr}1,\,\text{зад}}} \sum\limits_{j\,=\,y_{\text{поч}\,z,\,m}}^{y_{\text{поч}\,z,\,m}+h_{\mathbf{fr}1,\,\text{зад}}} (c^d_{1,\,\text{зад},\,(i,\,j)} - M_{1,\,\text{зад}})^2 \sum\limits_{i\,=\,x_{\text{поч}\,z,\,m}}^{2x_{\text{поч}\,z,\,m}+l_{\mathbf{fr}1,\,\text{зад}}} \sum\limits_{j\,=\,y_{\text{поч}\,z,\,m}}^{y_{\text{поч}\,z,\,m}+h_{\mathbf{fr}1,\,\text{зад}}} (c^d_{z,\,m,\,(i,\,j)} - M_{z,\,m})^2}}; \\ m = 1 \ldots N_{\chi_z};\ z = 2 \ldots N, \qquad (20)$$

де $c^d_{1,\,\text{зад},\,(i,\,j)}$ – значення кольору фрагмента $P_{1,\,\text{зад}}$; N_{χ_z} – розмірність покриттів $\chi_z \in \chi_{\left(\chi_{\mathbf{P}'}|(N-1)N_\chi\right),\, d_{D,\,\mathbf{fr}}}^{\sim\varepsilon/}$ та $\vartheta_z \in \vartheta_{\left(\vartheta_{\mathbf{P}'}|(N-1)N_\chi\right),\, d_{D,\,\mathbf{fr}}}^{\sim\varepsilon/}$.

В результаті (20) для кожного $\chi_{\left(\chi_{\mathbf{P}'}|(N-1)N_\chi\right),\, d_{D,\,\mathbf{fr}}}^{\sim\varepsilon/}$ та $\vartheta_{\left(\vartheta_{\mathbf{P}'}|(N-1)N_\chi\right),\, d_{D,\,\mathbf{fr}}}^{\sim\varepsilon/}$ отримуємо набір значень кореляцій $r_{z,\,m}(P_{1,\,\text{зад}}, P_{z,\,m})$, які є характеристиками фрагментів $P_{z,\,m} \in \vartheta_z$

$$\begin{pmatrix} \chi_z \\ \vartheta_z \end{pmatrix} \to \{r_{z,\,m}(P_{1,\,\text{зад}}, P_{z,\,m})\}_{m\,=\,1 \ldots N_{\chi_z}},\ z = 2 \ldots N; \\ \chi_z \in \chi_{\left(\chi_{\mathbf{P}'}|(N-1)N_\chi\right),\, d_{D,\,\mathbf{fr}}}^{\sim\varepsilon/}; \\ \vartheta_z \in \vartheta_{\left(\vartheta_{\mathbf{P}'}|(N-1)N_\chi\right),\, d_{D,\,\mathbf{fr}}}^{\sim\varepsilon/}. \qquad (21)$$

На наборі $\{r_{z,\,m}(P_{1,\,\text{зад}}, P_{z,\,m})\}$ для кожного z вирішуємо задачу пошуку кореляційного максимуму із заданим $P_{1,\,\text{зад}}$

$$I_{r,\,\max} = \left\{ \max_m (r_{z,\,m}(P_{1,\,\text{зад}}, P_{z,\,m})) \neq 0 \right\}_{z\,=\,2 \ldots N}. \qquad (22)$$

У випадку, якщо ненульового кореляційного максимуму при заданому z не існує, то це зображення видаляється з набору і в подальшому розв'язанні задачі суміщення не розглядається. Надалі вважатимемо, що для будь-якого z ненульова кореляція існує.

За (21) знаходимо відповідний $P_{z,\,m} \in \vartheta_z \in \vartheta_{\left(\vartheta_{\mathbf{P}'}|(N-1)N_\chi\right),\, d_{D,\,\mathbf{fr}}}^{\sim\varepsilon/}$ і формуємо остаточний набір фрагментів

$$\vartheta_{\mathbf{P}'\max r} = \left\{ P_{z,\,m} \middle| (P_{z,\,m} \to \mathbf{I}_{r,\,\max}) \right\}_{z\,=\,2 \ldots N} \qquad (23)$$

і відповідний набір фреймів

$$\chi_{\mathbf{P}'\max r} = \left\{ \mathbf{X}^{2,+,d}_{\mathbf{fr}\,z,\,m} \middle| P_{z,\,m} = C(\mathbf{X}^{2,+,d}_{\mathbf{fr}\,z,\,m}),\ P_{z,\,m} \in \vartheta_{\mathbf{P}'\max r} \right\}_{z\,=\,2 \ldots N}. \qquad (24)$$

Оскільки розмірність набору (22) дорівнює $N-2$, то

$$\dim\chi_{\mathbf{P}'\max r} = \dim\vartheta_{\mathbf{P}'\max r} = N-2. \qquad (25)$$

Очевидно, що $\vartheta_{\mathbf{P}'\max r} \subset \vartheta_{\left(\vartheta_{\mathbf{P}'}|(N-1)N_\chi\right),\, d_{D,\,\mathbf{fr}}}^{\sim\varepsilon/} \subseteq \subseteq \vartheta_{\mathbf{P}'} \subseteq \vartheta_{\mathbf{P}}$ належить топологіям $\mathfrak{V}_{\mathbf{P}'}$ та $\mathfrak{V}_{\mathbf{P}}$. Аналогічно для координатної області маємо $\chi_{\mathbf{P}'\max r} \subset \subset \chi_{\left(\chi_{\mathbf{P}'}|(N-1)N_\chi\right),\, d_{D,\,\mathbf{fr}}}^{\sim\varepsilon/} \subseteq \chi_{\mathbf{P}'} \subseteq \chi_{\mathbf{P}}$.

Фрагментний набір $(\vartheta_{\mathbf{P}'\max r}|N-2)$ є результатом двоетапного звуження $\vartheta_{\mathbf{P}'}$ до $N-2$ фрагментів, кожен з який відповідає окремому P_z набору $\mathbf{P}'$.

Подібно до $(\vartheta_{\mathbf{P}'\max r}|N-2)$, фреймовий набір $(\chi_{\mathbf{P}'\max r}|N-2)$ є результатом звуження $\chi_{\mathbf{P}'}$ і містить для кожного z по одному фрейму $\mathbf{X}^{2,+,d}_{\mathbf{fr}\,z,\,m}$.

За фреймовим набором $(\chi_{\mathbf{P}'\max r}|N-2)$, як зміщення між фрагментами $P_{z,\,m}$ і $P_{1,\,\text{зад}}$, знаходимо зміщення по осях $x - \Delta_{x,\,(1,\,z)}$ та по $y - \Delta_{y,\,(1,\,z)}$ кожного зображення набору $\mathbf{P}'$ відносно P_1,

$$\left.\begin{matrix}\Delta_{x,\,(1,\,z)} = \Delta_{x,\,z,\,m} - \Delta_{x,\,1,\,\text{зад}} \\ \Delta_{y,\,(1,\,z)} = \Delta_{y,\,z,\,m} - \Delta_{y,\,1,\,\text{зад}}\end{matrix}\right\}_{z\,=\,2\ldots N};$$

$$\mathbf{X}^{2,\,+,\,d}_{\mathbf{fr}z,\,m} \in \chi_{\mathbf{P}'\max r}. \qquad (26)$$

Звертаємо увагу на те, що зміщення $\Delta_{x,\,(1,\,z)}, \Delta_{y,\,(1,\,z)} \in \mathbf{N}$, тобто можуть набувати як додатних, так і від'ємних значень.

4. РЕЗУЛЬТАТИ ПРАКТИЧНИХ ЕКСПЕРИМЕНТІВ СУМІЩЕННЯ ЗОБРАЖЕНЬ НА ОСНОВІ ДИСПЕРСІЇ

На основі викладених вище теоретичних результатів розроблено практичну реалізацію методу суміщення зображень набору на основі дисперсії. Зображення цих наборів є результатами горизонтальних та вертикальних зсувів деякого базового зображення. Надалі такі набори будемо називати наборами штучно-згенерованих зображень (НШЗЗ).

В практичному експерименті для зручності умову (15) замінимо на відносну похибку

$$\frac{d_{D,\,\mathbf{fr}}(P_{z,\,m},\,P_{0,\,\text{зад}})}{|D_{z,\,m}|} \le \varepsilon. \qquad (27)$$

На рис. 1 наведено результати суміщення зображень НШЗЗ. Характеристики НШЗЗ є такими: розмірність набору – $N = 88$; зображення в градаціях сірого; розмірність кожного зображення – $l = 34 \times h = 54$ пікселів; $P_{\text{фікс}} = P_0$. Параметри заданого фрейму $\mathbf{X}^{2,\,+,\,d}_{\mathbf{fr}z,\,m}$: $\Delta_{x,\,0,\,\text{зад}} = \Delta_{y,\,0,\,\text{зад}} = 10$; $l_{\mathbf{fr}0,\,\text{зад}} = h_{\mathbf{fr}0,\,\text{зад}} = 10$; $\varepsilon = 0{,}001$. Індексування зображень в наборі розпочинається з нуля, тобто $\mathbf{P}' = \{P_1, \ldots, P_{88}\}$. Заданий фрагмент на P_0 виділений червоним кольором.

На рис. 2 наведено результати побудови майже фактор простору $\vartheta^{\,/\,\sim\varepsilon}_{\left(\vartheta_{\mathbf{P}'}|(N-1)N_\chi\right),\,d_{D,\,\mathbf{fr}}}$, тобто набори «підозрілих на подібність» фрагментів для кожного зображення набору $\mathbf{P}'$, НШЗЗ, результати суміщення якого наведені на рис. 1. Швидкість формування $\vartheta^{\,/\,\sim\varepsilon}_{\left(\vartheta_{\mathbf{P}'}|(N-1)N_\chi\right),\,d_{D,\,\mathbf{fr}}}$ є визначальною для пропонованого алгоритму в порівнянні з відомими методами [7]. Чисельні значення кількості «підозрілих фреймів» для кожного P_z набору $\mathbf{P}'$ можна побачити на рис. 1 в таблиці «Зміщення» → в колонці «значення» → в мітці «підозр. –».

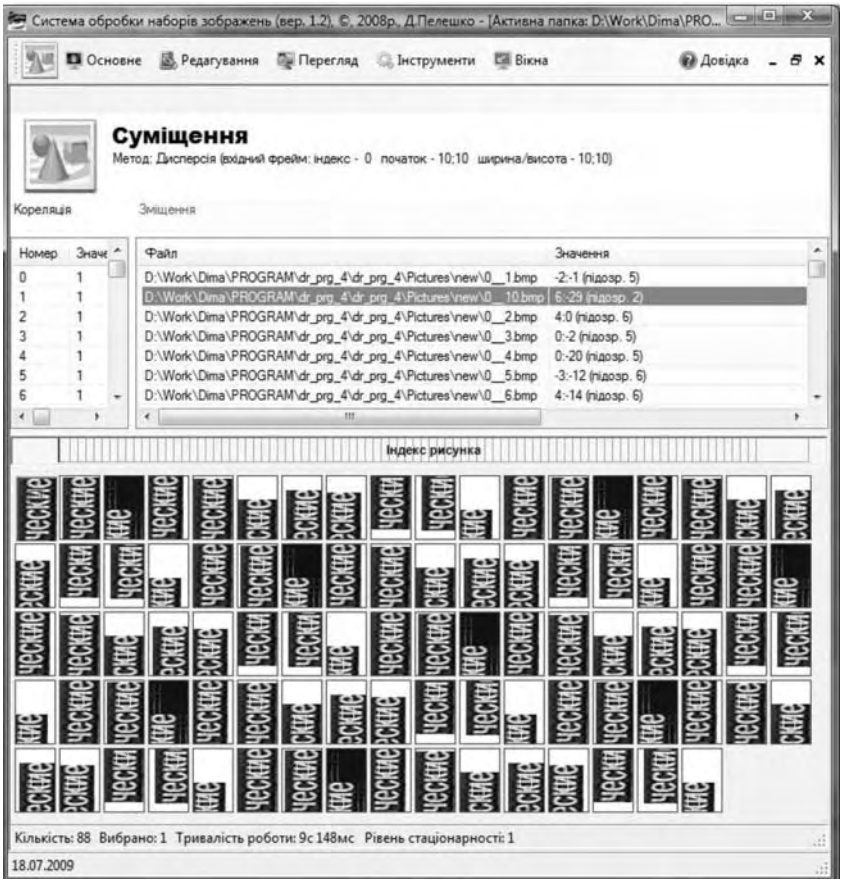


Рис. 1. Зріз екранної форми – результати суміщення на основі дисперсії НШЗЗ

Рис. 2. Зріз екрану – результати формування майже фактор простору $\vartheta_{\left(\vartheta_{\mathbf{P}'}|(N-1)N_{\chi}\right),\, d_{D,\,\mathbf{fr}}}{}^{/\sim\varepsilon}$ при суміщенні методом дисперсії кольору НШЗЗ $\mathbf{P}'$

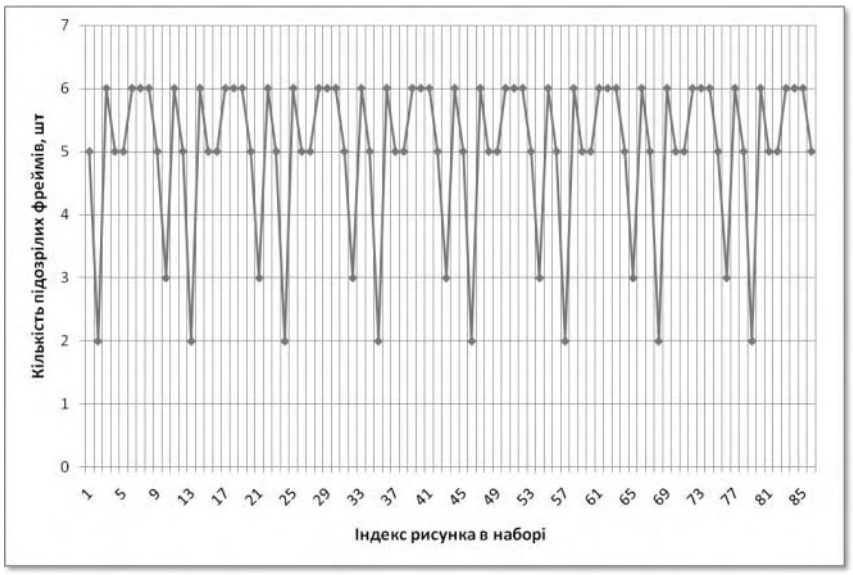


Рис. 3. Розподіл «підозрілих» фрагментів (простору $\vartheta_{\left(\vartheta_{\mathbf{P}'}|(N-1)N_{\chi}\right),\, d_{D,\,\mathbf{fr}}}{}^{/\sim\varepsilon}$) в невпорядкованому НШЗЗ $\mathbf{P}'$ за методом суміщення на основі дисперсії

На рис. 3 наводиться розподіл розмірностей χ_z та ϑ_z, які належать покриттям $\chi_{\left(\chi_{\mathbf{P}'}|(N-1)N_{\chi}\right),\, d_{D,\,\mathbf{fr}}}{}^{/\sim\varepsilon}$ та $\vartheta_{\left(\vartheta_{\mathbf{P}'}|(N-1)N_{\chi}\right),\, d_{D,\,\mathbf{fr}}}{}^{/\sim\varepsilon}$ відповідно. Періодичність розподілу визначається невпорядкованістю набору $\mathbf{P}'$ і штучним генеруванням зображень.

Чисельні значення, тобто $\Delta_{x,\,(1,\,z)}, \Delta_{y,\,(1,\,z)}$, для суміщення на основі дисперсії наведені на рис. 1 в таблиці «Зміщення» → в колонці «значення».

На рис. 4 наведено часові результати (тобто, фактично швидкість) роботи алгоритму суміщення НШЗЗ запропонованим методом залежно від розмірності $\mathbf{P}'$. Характеристики НШЗЗ є такими: розмірність набору – змінна; зображення в градаціях сірого; розмірність кожного зображення – $l = 34 \times h =$ = 54 пікселів; $P_{\text{фікс}} = P_0$. Параметри заданого фрейму $\mathbf{X}^{2,\,+,\,d}_{\mathbf{fr}0,\,\text{зад}}$: $\Delta_{x,\,0,\,\text{зад}} = \Delta_{y,\,0,\,\text{зад}} = 10$; $l_{\mathbf{fr}0,\,\text{зад}} = h_{\mathbf{fr}0,\,\text{зад}} =$ = 10; ε = 0,01.

При суміщенні методом дисперсії результати наведено в табл. 1. При цьому середньоквадратичне відхилення результатів експерименту не перевищувало 20 мс. На основі результатів вибраних експериментів характеристикою швидкості роботи алгоритму виступало середнє значення значень часу усіх експериментів при кожній розмірності набору. Ці значення наведені на графіку рис. 4.

З рис. 4 видно, що час роботи алгоритму зростає із збільшенням розмірності набору. Таке зростання пояснюється різким збільшенням арифметичних операцій.

Таблиця 1. Зведена таблиця експериментальних та характеристичних даних – результатів роботи (в мс) процесу суміщення методом дисперсії при різних розмірностях НШЗЗ

Розмірність набору	Час роботи алгоритму, мс Номер експерименту					Відхилення, мс	Середнє значення, мс
	1	2	3	4	5		
11	1104	1089	1102	1096	1098	5,85	1097,80
22	2272	2248	2263	2247	2238	13,65	2253,60
32	3283	3301	3258	3293	3294	16,81	3285,80
44	4607	4624	4593	4613	4610	11,19	4609,40
88	9148	9145	9112	9107	9112	19,94	9124,80

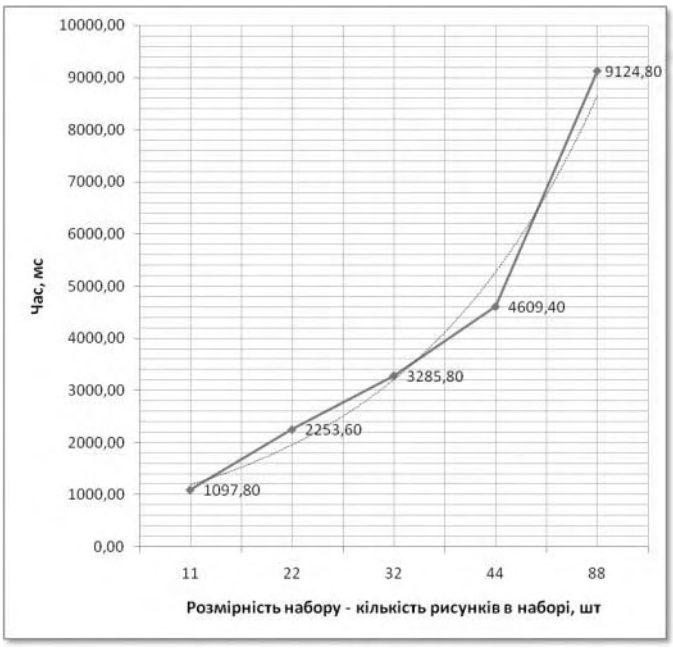


Рис. 4. Часова залежність від розмірності набору $\mathbf{P}'$ роботи алгоритму суміщення НШЗЗ методом дисперсії

Таблиця 2. Порівняльні дані результатів роботи (в мс) процесу суміщення НШЗЗ методом дисперсії при різних розмірах заданого фрейма

Розмір рисунка (пікселі)			
X		Y	
37		54	
Розмір фрейма (пікселі) X	Y	Площа фрейма / площа рисунка	Час
10	10	0,05	1097,80
15	15	0,11	1491,20
20	20	0,20	1772,40
25	25	0,31	1558,20
28	28	0,39	1222,20

В табл. 2 наведено дані залежності швидкості роботи процесу суміщення НШЗЗ від розмірів фрейма $\mathbf{X}_{\mathbf{fr}0,\,\text{зад}}^{2,\,+,\,d}$. Характеристики НШЗЗ є такими, як у випадку з результатами, поданими в табл. 2.

Експерименти проводились подібно до експериментів, результати яких відображені на рис. 4 і в табл. 1. Тобто характеристикою виступало середнє значення результатів п'яти кращих експериментів при різних $s_{0,\,\text{зад}}/s_{P_0}$. При цьому похибка відхилення також не перевищувала 20 мс. Як показали результати експериментів, найгіршим (найдовше працював алгоритм) для даного $\mathbf{P}'$ є співвідношення $s_{0,\,\text{зад}}/s_{P_0} = 0,2$, що відповідає розмірам $l_{\mathbf{fr}0,\,\text{зад}} = h_{\mathbf{fr}0,\,\text{зад}} = 20$.

При більших та менших розмірах заданого фрейма (в даному випадку квадратного) швидкість роботи алгоритму лише зростає.

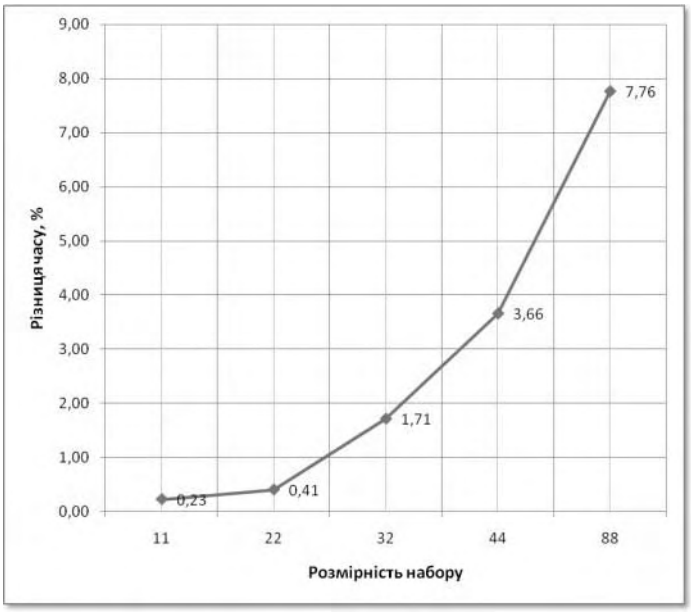


Рис. 5. Порівняння часу роботи алгоритму суміщення НШЗЗ на основі методів кореляційної прив'язки і дисперсії

Наведені результати ілюструють лише характер залежності швидкості роботи алгоритму від розмірів $\mathbf{X}_{\mathbf{fr}0,\,\text{зад}}^{2,\,+,\,d}$ (тренд поліномний). Очевидно, що тип і якісний вміст зображення та вибраного фрагмента також будуть впливати на час роботи алгоритму.

Важливим результатом експериментів є існування максимуму – найбільшого часу роботи алгоритму. Відповідно до цього можна зробити висновок, що пришвидшення роботи алгоритму є можливим через вибір за розмірами $\mathbf{X}_{\mathbf{fr}0,\,\text{зад}}^{2,\,+,\,d}$. Пошук найменшого значення є достатньо складним, оскільки до розгляду треба приймати двомірний розподіл часу роботи.

На рис. 5 показано порівняння часу роботи різних алгоритмів (у форматі приросту у відсотках пришвидшення роботи запропонованого методу в порівнянні з методом кореляційної прив'язки для суміщення НШЗЗ) суміщення, побудованих на методах кореляційної прив'язки та дисперсії.

ВИСНОВКИ

Як можна побачити з результатів, наведених на рис. 5, метод суміщення, базований на майже факторизації простору $\vartheta_{\mathbf{P}'}$ на основі дисперсії, є суттєво швидшим від методу суміщення на основі кореляційної прив'язки. Зважаючи на дуже малі розміри зображень НШЗЗ, приріст швидкості роботи, наприклад при N = 88 становить 7,76 % і зростає при зростанні розмірності $\mathbf{P}'$.

Запропонований алгоритм може бути застосований для суміщення в горизонтальному та вертикальному напрямках зображень будь-якого типу.

СПИСОК ЛІТЕРАТУРИ

1. *Гусейн-Заде, С. М.* Лекции по дифференциальной геометрии / Гусейн-Заде С. М. – М. : Изд-во МГУ, 2001.– 464 с.
2. *Милнор, Дж.* Дифференциальная топология / Дж. Милнор, А. М. Уоллес. – М. : Мир, 1972. – 279 с.
3. Класифікація моделей представлення зображень та наборів зображень як стохастичних зображень та полів: Матеріали науково-практичної конференції [«Інтелектуальні системи прийняття рішень та проблеми обчислювального інтелекту ISDMCI'2009»], (Євпаторія, 18–22 травня 2009) / Херсонський морський інститут. – Херсон : Видавництво Херсонського морського інституту, 2009. – Т. 2. – С. 401–405.
4. *Пелешко, Д. Д.* Топології зображень та наборів зображень / Д. Пелешко // Науковий вісник НЛТУ України : збірник науково-технічних праць. – 2009. – Вип. 19.4. – С. 236–242.
5. *Александров, П. С.* Введение в теорию множеств и общую топологию / Александров П. С. – М. : Наука, 1977. – 368 с.
6. *Халмош, П.* Конечномерные векторные пространства / Халмош П. – М. : ГИФМЛ, 1963. – 276 с.
7. *Рашкевич, Ю.* Центрування зображень на основі методів кореляційного аналізу / Ю. Рашкевич, Б. Демида, Д. Пелешко, Н. Кустра // Збірник наукових праць Інститут проблем моделювання в енергетиці ім. Г. Є. Пухова. – 2005. – Вип. 29. – С. 121–128.

Надійшла 7.10.2009
Після доробки 30.03.2010

Пелешко Д. Д., Шпак З. Я., Кустра Н. Я.

СОВМЕЩЕНИЕ ИЗОБРАЖЕНИЙ НАБОРА НА ОСНОВЕ ИСПОЛЬЗОВАНИЯ ДИСПЕРСИИ ЦВЕТА ИЗОБРАЖЕНИЙ

Разработан ускоренный метод центрирования набора однотипных изображений на основе решения задачи почти

факторизации пространства топологиии изображения с последующим сужением этого пространства с помощью задачи поиска корреляционного максимума. Задача почти факторизации формулируется введением полуметрики относительно дисперсии цвета элементов топологии изображения.

Ключевые слова: совмещение изображений, фреймовое покрытие, топология изображений, дисперсия цвета, корреляционный максимум.

Peleshko D. D., Kustra N. O., Shpak Z. Ya.

COMPOSITION IMAGE REGISTRATION USING PICTURE COLOR DISPERSION

The authors have developed the method of one-type images centering based on solution of the problem of image topology space almost-factorization with further constriction of the space by solving the problem of correlation maximum search. The almost-factorization problem is solved by introduction of semimetrisc relative to image topology elements color dispersion.

Key words: image registration, frame coverage, image topology, color dispersion, correlation maximum.

УДК 629.735

Потий А. В.[1], Комин Д. С.[2]

[1]Д-р техн. наук, доцент, начальник кафедры Харьковского университета Воздушных Сил

[2]Адъюнкт Харьковского университета Воздушных Сил

ОНТОЛОГИЧЕСКОЕ МОДЕЛИРОВАНИЕ ПРОЦЕССА ОЦЕНИВАНИЯ ГАРАНТИЙ В КОНТЕКСТЕ ФУНКЦИОНАЛЬНО-ЛИНГВИСТИЧЕСКОГО ПОДХОДА

Предлагается функционально-лингвистический подход к оцениванию гарантий безопасности. Приводятся методика и результаты онтологического анализа предметной области оценивания гарантий. Обосновывается актуальность применения аппарата онтологического моделирования для задач оценивания безопасности.

Ключевые слова: гарантии, оценивание, онтологическое моделирование, функциональное моделирование, лингвистические переменные.

ВСТУПЛЕНИЕ

Владельцы систем и продуктов информационных технологий (ИТ) хотят быть уверенными в качестве разработки, эффективности функционирования и безопасности приобретенных ИТ-продуктов. Международные [1–4] и национальные [5–7] стандарты в области безопасности информационных технологий определяют функциональные требования безопасности и требования гарантий безопасности, удовлетворение которых позволяет предоставить различные основания для такой уверенности. В ходе активного исследования (оценивания) ИТ-продукта на соответствие требованиям гарантий и формируется уверенность потребителя в корректности реализации функциональных услуг безопасности.

Оценивание ИТ-продуктов проводится аккредитованными испытательными лабораториями на основании программ и методик проведения оценивания. Качественная разработка программы и методики оценивания является важной составляющей при подготовке к проведению оценивания. Сам процесс оценивания подвержен воздействию различных факторов, способных повлиять на итоговый результат оценивания. Поэтому к процессу оценивания выдвигаются требования ширины, глубины и строгости, а к результатам оценивания – требования объективности, повторяемости, беспристрастности, воспроизводимости и сопоставимости.

Обзор научной литературы показал, что моделирование процессов оценивания гарантий безопасности с созданием инструментальных средств для поддержки работы эксперта является актуальной задачей. Однако в основном моделирование направлено на интерактивное представление требований стандарта в виде информационных инструментальных систем. Кроме того, в большинстве работ при моделировании не рассматриваются вышеперечисленные требования.

В работах авторов [8–10] предлагается функционально-лингвистический подход к оцениванию гарантий информационной безопасности, который позволяет разрабатывать программу и методику оценивания и выполнять вышеуказанные требования как к процессу оценивания, так и к результатам оценивания. В данной работе представлены результаты дальнейшего развития функционально-лингвистического подхода и детальное описание первого этапа.

1. КОНЦЕПЦИЯ ФУНКЦИОНАЛЬНО-ЛИНГВИСТИЧЕСКОГО ПОДХОДА

В ходе анализа области оценивания гарантий авторы выдвинули предложение оценивать не сам объект оценивания, а присущие ему *свойства гарантий*. Данные свойства выявляются в ходе анализа требований, выдвигаемых к объекту в нормативных документах на определенном уровне гарантий безопасности. Для доказательства того, что конкретное свойство присуще (имеется в наличии) объекту оценивания (ОО), используются свидетельства, в качестве которых могут выступать сам ОО, его части, документация, результаты испытаний (тестирования). Таким образом, оценивание ИТ-продукта заключается в оценке степени проявления присущих продукту свойств гарантий на основании анализа (исследования) свидетельств. Это является ключевой идеей предлагаемого авторами подхода к оцениванию гарантий безопасности.

Структура функционально-лингвистического подхода представлена на рис. 1. Оценка гарантий безопасности осуществляется в четыре этапа.

На первом этапе проводится онтологический анализ и моделирование предметной области оценивания. Анализ включает в себя исследование множества требований гарантий ($R = \{r_1, r_2, \ldots, r_i\}$, $i = \overline{1, N}$), которые выдвигаются к ОО, и выявления (синтез) множества свойств гарантий ($P = \{p_1, p_2, \ldots, p_j\}$, $j = \overline{1, L}$), которыми должен обладать ОО. На множестве свойств гарантий P определяются зависимости и взаимосвязи между свойствами. Результаты анализа представляются в виде онтологических графов,

Рис. 1. Функционально-лингвистический подход к оцениванию гарантий

которые точно и однозначно (в принятой нотации) описывают предметную область (т. е. основные понятия (концепты) и отношения между ними). Полнота охвата моделируемой области оценивания гарантий обеспечивается использованием онтологических графов двух видов: объектно-ориентированных и процессно-ориентированных онтологий.

На втором этапе осуществляется функциональное моделирование процесса оценивания гарантий. Целью функционального моделирования является формализованное представление процесса оценивания. В качестве языка моделирования была выбрана нотация IDEF0 [8]. Нотация IDEF0 позволяет однозначно отобразить шаги проведения оценивания (в виде направленного графа), для каждого шага определить оцениваемое свойство гарантий, свидетельства, необходимые для оценивания данного свойства гарантий, субъектов оценивания и нормативное основание проведения оценивания. Если оценке подлежит сложное (составное) свойство гарантий, то каждый шаг (блок IDEF0-диаграммы) может быть декомпозирован для детального описания процедуры оценки подсвойств.

На третьем этапе для каждого свойства p_j вводится лингвистическая переменная $\Omega p_j = <\beta, T(\beta), G, M>$ и определяется ее терм-множество β. Применение математического аппарата лингвистических переменных обусловлено тем, что чаще всего степень проявления свойства гарантий не может быть описана с помощью количественных показателей. Поэтому для принятия решения относительно степени проявления свойства гарантий удобно использовать математический аппарат нечеткого логического вывода (НЛВ) на основе заранее подготовленной базы продукционных правил [11]. Применение лингвистических переменных и операций нечеткого логического вывода обеспечивает выполнение требований объективности и повторяемости результатов оценивания гарантий.

На четвертом этапе строятся диаграммы потоков работ в нотации IDEF3. Это позволяет однозначно установить порядок и приоритетность выполнения операций (действий) оценивания. Каждый блок диаграммы представляет собой отдельное действие эксперта-оценщика. После каждого блока следует перекресток, который определяет критерий (правило) выбора следующего действия в зависимости от того, какое решение примет эксперт относительно степени проявления оцениваемого свойства. Количество вариантов выбора зависит от количества значений, которые может принимать лингвистическая переменная, описывающая оцениваемое свойство. Диаграммы

определяют точки, в которых эксперт должен принять решение и вынести вердикт относительно степени проявления того или иного свойства. Построение и применение IDEF3 диаграмм обеспечивает выполнение требования повторяемости результатов оценивания, т. к. для каждой операции оценивания определяется набор вариантов возможных вердиктов (решений) эксперта. Выбор варианта вердикта зависит от того, какие значения принимают лингвистические переменные в ходе оценивания свойств (по сути это выбор эксперта относительно степени проявления свойства).

Таким образом, реализация данного подхода позволяет обеспечить выполнение основных требования к процессу и результатам оценивания гарантий безопасности.

2. ОНТОЛОГИЧЕСКИЙ АНАЛИЗ И МОДЕЛИРОВАНИЕ ПРЕДМЕТНОЙ ОБЛАСТИ ОЦЕНИВАНИЯ ГАРАНТИЙ

В ходе онтологического анализа требований гарантий необходимо найти консенсус (баланс) между степенью детализации оцениваемых свойств гарантий и стоимостью процесса оценивания. Чем больше глубина оценивания (выше степень детализации), тем выше точность оценки объекта оценки. Однако при этом увеличиваются стоимостные временные показатели процесса оценивания. Низкая детализация требований удешевляет процесс оценивания, однако может привести к затруднениям при принятии решении о степени проявления оцениваемых свойств и, как следствие, к неправильным выводам.

2.1. Задачи онтологического анализа и выбор типа онтологии

Основными задачами онтологического анализа требований гарантий являются:

– точное, ясное и однозначное описание предметной области оценивания гарантий информационной безопасности, выделение основных понятий и концептов;

– четкое определение содержания оценивания;

– определение необходимой глубины оценивания;

– преобразование набора статических требований гарантий в динамическую базу для их использования людьми из разных сфер деятельности и понимания области оценивания;

– представление знаний (требований) области оценивания гарантий в форме, которая позволяет создавать ее электронный аналог.

Анализ структуры требований гарантий [3], где ключевыми *объектами* структуры являются «класс – семейство – компонент – элемент», показал, что для моделирования требований необходимо использовать иерархические объектно-ориентированные графы (онтологии). Пример результата моделирования представлен на рис. 2.

В узлах графа находятся выделенные объекты (класс – семейство – компонент – элемент). Ребра графа показывают связь между объектами. Так, классы гарантий декомпозированы на семейства. Каждое

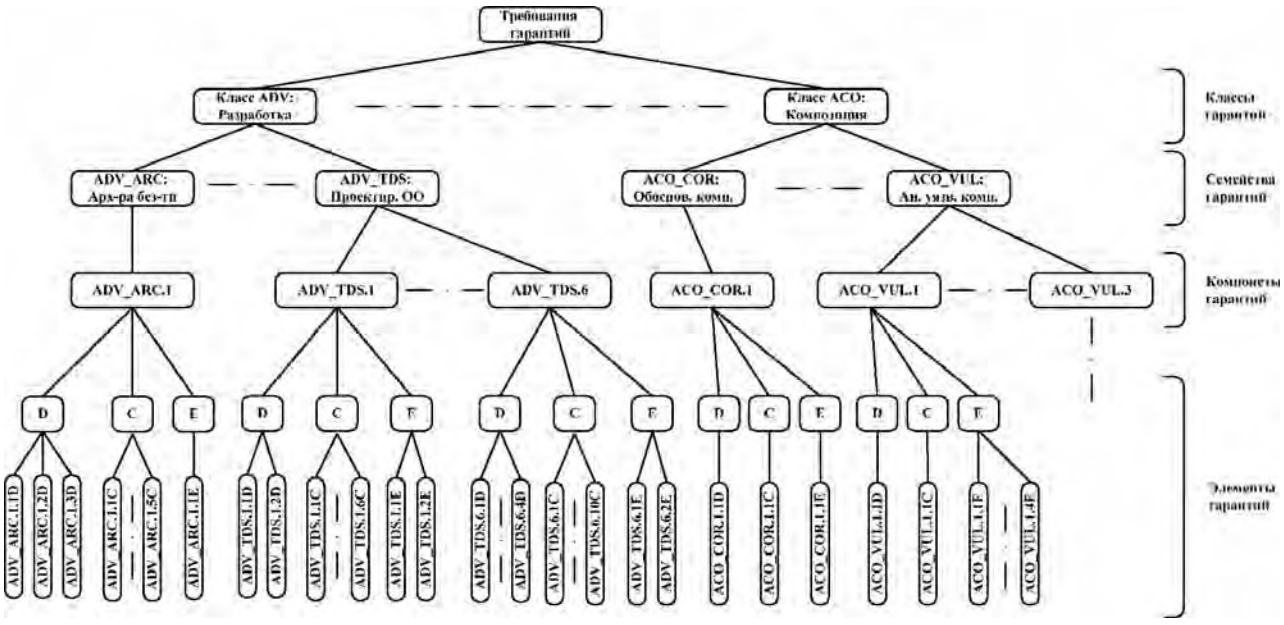


Рис. 2. Иерархическая объектно-ориентированная онтология требований гарантий

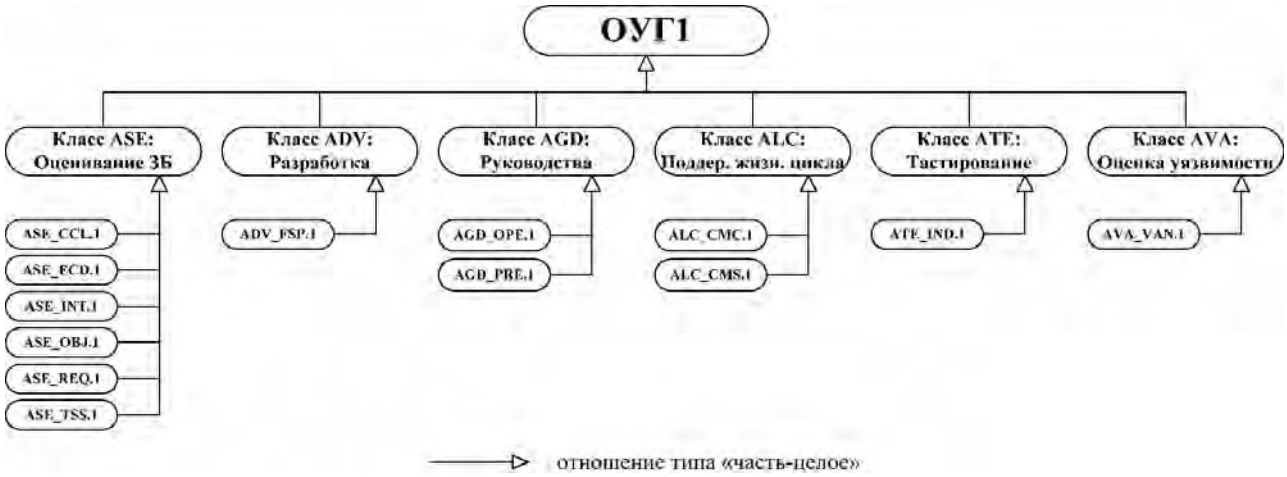


Рис. 3. Иерархическая объектно-ориентированная онтология требования гарантий по ОУГ1

семейство гарантий содержит набор атрибутов, предоставляющих информацию о характеристике, ранжировании компонентов и особенностях применения семейства. Каждое семейство гарантий содержит множество связанных с ним компонентов. При моделировании компоненты могут быть выражены как субсемейства. На уровне компонентов возможны отношения между компонентами как внутри одного класса, так и межклассовые. Каждый компонент содержит элементы гарантий трех типов: элементы действий разработчика, элементы содержания и представления свидетельств, элементы действий оценщика.

Пример иерархической объектно-ориентированной онтологии требований гарантий по оценочному уровню гарантий 1 (ОУГ1) представлен на рис. 3.

2.2. Онтологическая модель области оценивания гарантий

Онтологическая модель области оценивания гарантий включает в себя *объектно-ориентированную онтологическую модель* и *процессно-ориентированную модель* оценивания гарантий информационной безопасности.

2.2.1. Способ построения объектно-ориентированной онтологии области оценивания гарантий

Построение объектно-ориентированной онтологии области оценивания гарантий осуществляется в 3 этапа (рис. 4).

I этап. Строится объектно-ориентированный иерархический граф требований гарантий (G^R). Определяется степень глубины (уровень детализации) требований. Степень детализации требований определяет мощность множества свойств гарантий объекта оценки. Выявляются отношения зависимости на множестве требований гарантий и определяется их тип (часть – целое, экзистациональная, каузальная, внутриклассовая, межклассовая и т. д.). Формальная запись графа требований гарантий имеет вид

$$G^R = <R, Q_R>, \quad (1)$$

где $R = \{r_1, r_2, \ldots, r_i\}$, $i = \overline{1, N}$ – множество требований гарантий, $Q_R = \{Q_f[r_i \leftrightarrow r_j]\}$, $f = \overline{1, F}$ – множество отношений на множестве требований гарантий.

II этап. Строится объектно-ориентированный иерархический граф свойств гарантий (G^P), которые необходимо оценить. Между иерархиями требований и свойств гарантий устанавливаются отношения (зависимости) ($D[R \leftrightarrow P]$). На основе анализа зависимостей Q_R определяются зависимости свойств гарантий Q_P. Зависимости могут как повторяться, так и возникать новые. Определяются сложные свойства гарантий, т. е. такие свойства, для оценки которых необходимо проверить или исследовать множество субсвойств. Формальная запись графа свойств гарантий имеет вид

$$G^P = <P, Q_P>, \quad (2)$$

где $P = \{p_1, p_2, \ldots, p_i\}$, $i = \overline{1, N}$ – множество свойств гарантий, $Q_P = \{Q_s[p_i \leftrightarrow p_j]\}$, $s = \overline{1, S}$ – множество отношений между свойствами гарантий.

III этап. Строится иерархический граф множества свидетельств (G^E), которые получены путем декомпозиции объекта оценки. Для каждого элементарного свойства $p_i \in P$ определяется множество свидетельств $Ep_i = \{e_1, e_2, \ldots, e_i\}$, $i = \overline{1, N}$, которые необходимо исследовать для оценки данного свойства. За-

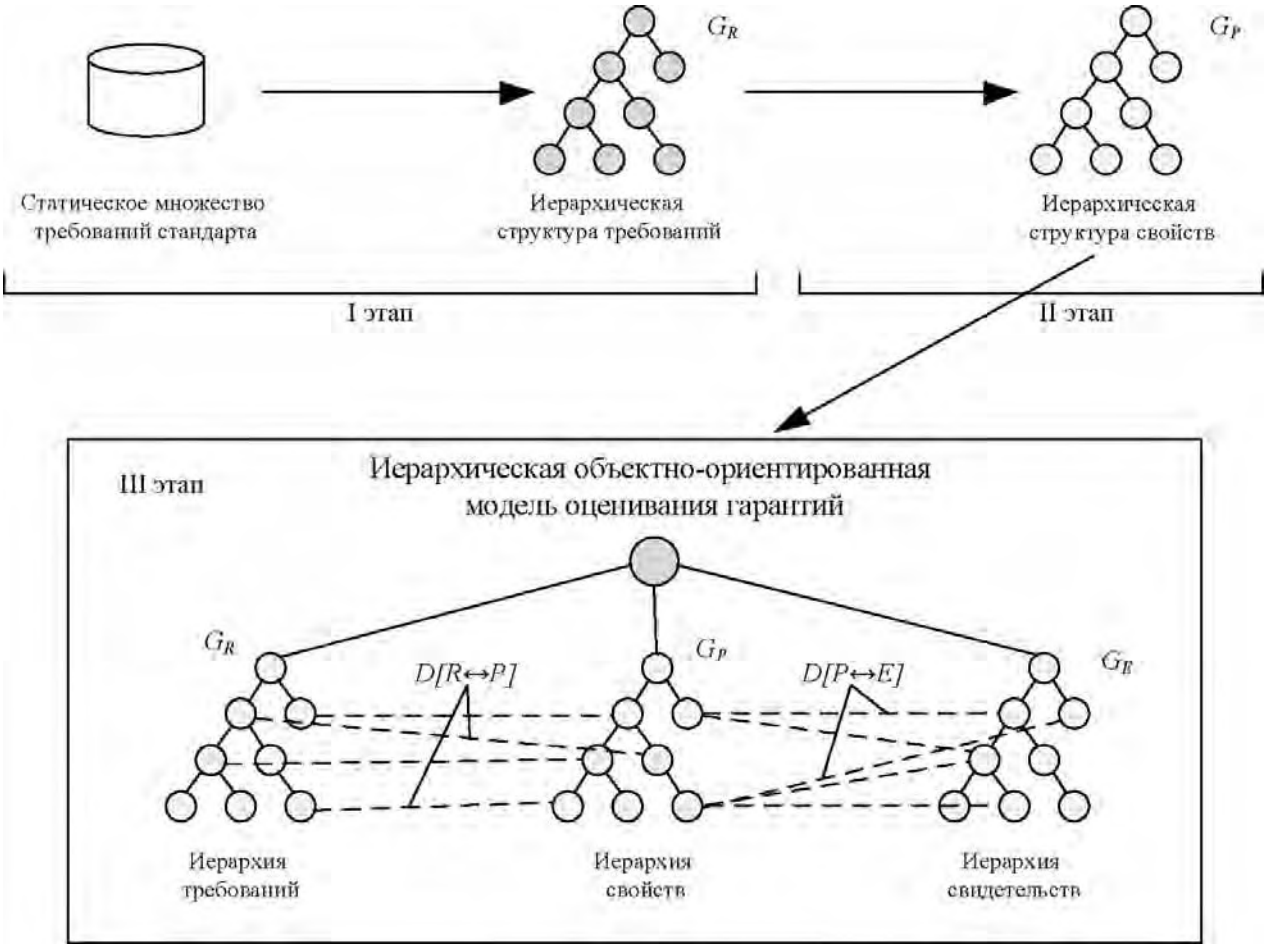


Рис. 4. Объектно-ориентированная модель области оценивания гарантий

висимости между графами G^P и G^E показываются в виде отношений типа «свойство – свидетельство» $D[PE]$. Формальная запись графа свидетельств имеет вид

$$G^E = <E, Q_E>, \quad (3)$$

где $E = \{e_1, e_2, \ldots, e_z\}$, $z = \overline{1, Z}$ – множество свидетельств, $Q_E = \{Q_y[e_i \leftrightarrow e_j]\}$, $y = \overline{1, Y}$ – множество отношений между свидетельствами.

Для сложных свойств соответствие между свидетельствами и свойствами может быть представлено в табличном (табл. 1) или матричном виде (4).

Таблица 1. Соответствие свидетельств и свойств

Сложное свойство, P			
e_i p_j	p_1	p_2	p_3
e_1	1	1	1
e_2	0	1	0
e_3	1	0	0

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}. \quad (4)$$

В ячейки таблицы соответствия ставится «1» тогда, когда свидетельству e_i (строки таблицы) присуще свойство p_j (столбцы таблицы), то есть когда для оценки свойства p_j, записанного в столбце, необходимо использовать свидетельство e_i. В ячейке таблицы записывается «0», если между свидетельством и свойством нет взаимосвязи (соответствия).

Заполнение таблицы соответствия упорядочивает (систематизирует) знания об объекте оценки и позволяет уточнять характеристику оцениваемых свойств. Это в конечном итоге оказывает влияние на объективность оценки свойств гарантий.

Мощность множества требований (свойств, свидетельств), отображаемых с использованием онтологий иерархического типа, может быть определена по формуле

$$W = \sum_i \sum_h \sum_l G^i \cdot S_{h,\,l}, \qquad (5)$$

где G^i – онтограф i-го множества, $i = \overline{1, 3}$; $S_{h,l}$ – степень вершины, равная числу исходящих из нее ребер, $h = \overline{1, H}$ – количество уровней онтологического графа, $l = \overline{1, L_h}$ – номер вершины на соответствующем (h-м) уровне онтологического графа.

Формально, объектно-ориентированная онтология области оценивания гарантий Ω_{ASO} имеет вид

$$\Omega_O = <G^R, G^P, G^E, D>, \qquad (6)$$

где G^R – объектно-ориентированный онтологический граф множества требований гарантий; G^P – объектно-ориентированный онтологический граф множества свойств; G^E – объектно-ориентированный онтологический граф множества свидетельств; $D = \{D[R \leftrightarrow P], D[P \leftrightarrow E]\}$ – множество отношений типа «требование – свойство» и «свойство – свидетельство».

Таким образом, для каждого свойства гарантий однозначно определяется требование, с которым оно связано, и свидетельство (одно или множество), необходимое для оценки данного свойства.

Пример объектно-ориентированной онтологии по оцениванию документации, которая описывает процедуры установки, генерации и запуска (ПУГЗ), представлен на рис. 5.

2.2.2. Способ построения процессно-ориентированной онтологии области оценивания гарантий

Процессно-ориентированная онтология оценивания гарантий строится на основе требований международного стандарта ISO/IEC 18045 [4]. Целью ее построения является определение взаимосвязи между оцениваемыми свойствами и действиями по оценке гарантий безопасности. Построение процессно-ориентированной онтологии осуществляется с учетом результатов, полученных при построении объектно-ориентированной онтологической модели. В качестве входов для выполнения процедуры построения процессно-ориентированой онтологии выступают, в частности, онтологический граф требований G^R, онтологический граф свойств гарантий G^P и множество отношений между ними $D[R \leftrightarrow P]$.

I этап. Строится онтологический граф действий по оценке гарантий G^A, определенных в стандарте ISO/IEC 18045 [4]. Определяется множество отношений зависимости Q_A на множестве действий A. Формальная запись графа действий имеет вид

$$G^A = <A, Q_A>, \qquad (7)$$

где $A = \{a_1, a_2, \ldots, a_i\}$, $i = \overline{1, N}$ – множество действий по оценке гарантий, $Q_A = \{Q_s[a_i \leftrightarrow a_j]\}$, $s = \overline{1, S}$ – множество отношений между действиями по оценке гарантий.

II этап. Определяется множество зависимостей $D[R \leftrightarrow A]$ между онтологическими графами действий G^A и требований G^R. Взаимосвязь между структурными компонентами требований гарантий и структурой действий по стандарту ISO/IEC 18045 представлена на рис. 6.

III этап. Определяется множество зависимостей $D[A \leftrightarrow P]$ между онтологическими графами действий по оценке гарантий G^A и свойств гарантий G^P. Т. к. напрямую данные зависимости определить нельзя, они определяются косвенным путем, т. е. через требования гарантий (рис. 7).

IV этап. Строится онтология субъектов G^B, вовлеченных в процесс оценивания гарантий информационной безопасности. Формальная запись графа заинтересованных субъектов имеет вид

$$G^B = <B, Q_B>, \qquad (8)$$

где $B = \{b_1, b_2, \ldots, b_i\}$, $i = \overline{1, N}$ – множество заинтересованных субъектов, $Q_B = \{Q_f[b_i \leftrightarrow b_j]\}$, $f = \overline{1, F}$ – множество взаимосвязей между субъектами.

Онтология субъектов, которые принимают участие в процессе обеспечения и оценивания гарантий безопасности, представлена на рис. 8 [13].

Таким образом, формальная запись процессно-ориентированной онтологии оценивания гарантий безопасности Ω_{ASP} имеет вид

$$\Omega_P = <G^R, G^P, G^A, D, G^B>, \qquad (9)$$

где G^R – онтологический граф множества требований гарантий; G^P – онтологический граф множества свойств; G^A – онтологический граф множества действий по оценке гарантий; $D = \{D[R \leftrightarrow P], D[R \leftrightarrow A], D[A \leftrightarrow P]\}$ – множество отношений типа «требование – свойство», «требование – действие» и «действие – свойство»; G^B – онтологический граф субъектов, вовлеченных в процесс оценивания гарантий.

Пример процессно-ориентированной онтологии по оцениванию документации, которая описывает ПУГЗ, представлен на рис. 9.

ЗАКЛЮЧЕНИЕ

Исследование требований гарантий с использованием аппарата онтологического моделирования дает более глубокое понимание предметной области оце-

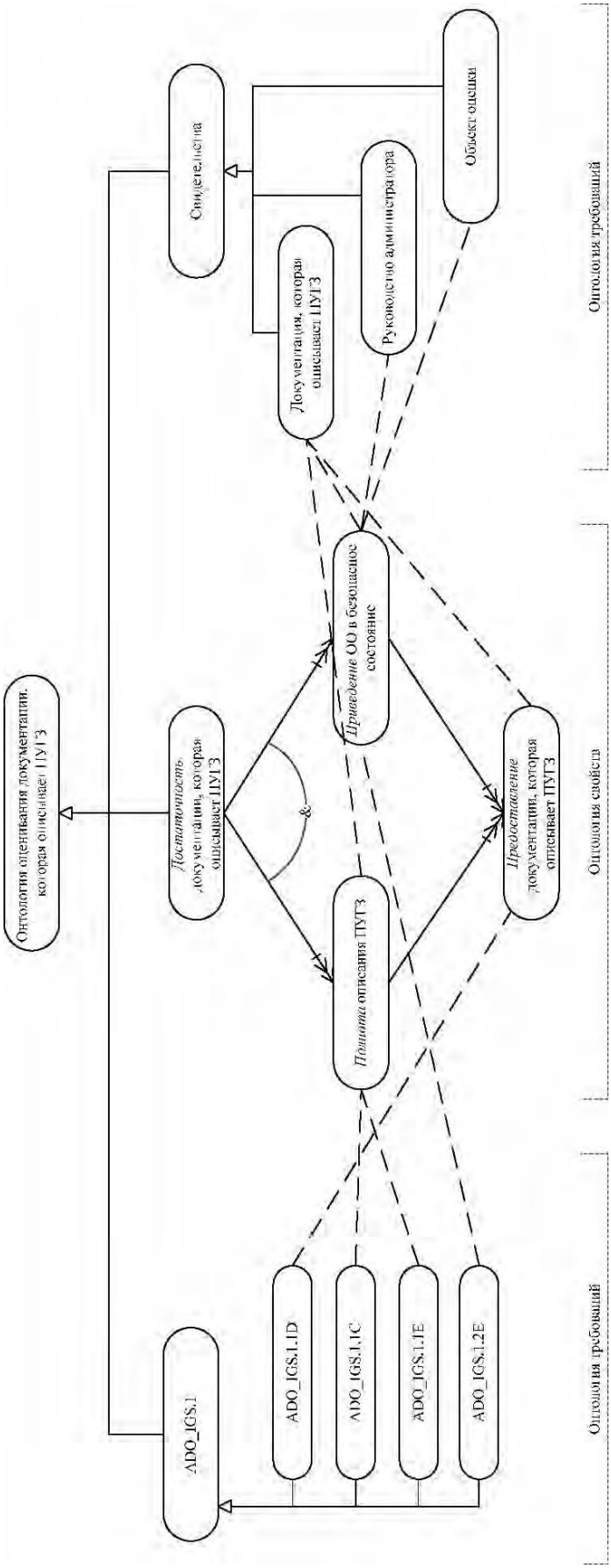


Рис. 5. Объектно-ориентированная онтология по оцениванию документации, которая описывает ПУГЗ

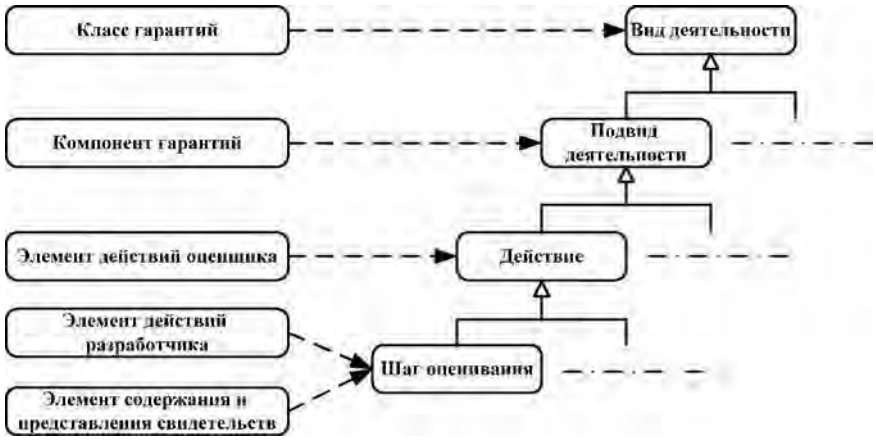


Рис. 6. Соответствие между компонентами требований гарантий и действиями по их оценке

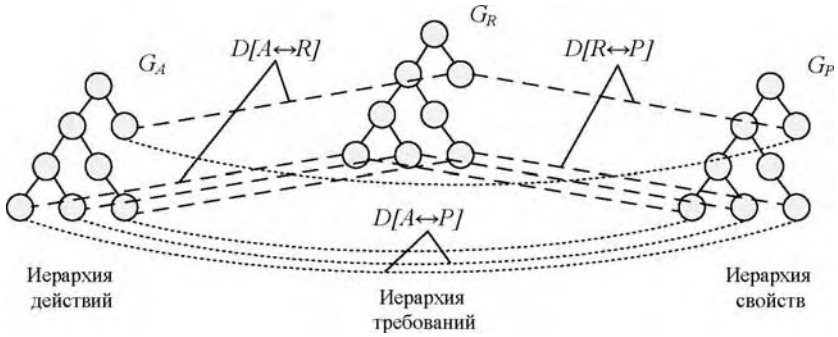


Рис. 7. Модель соответствия действий по оценке и свойств

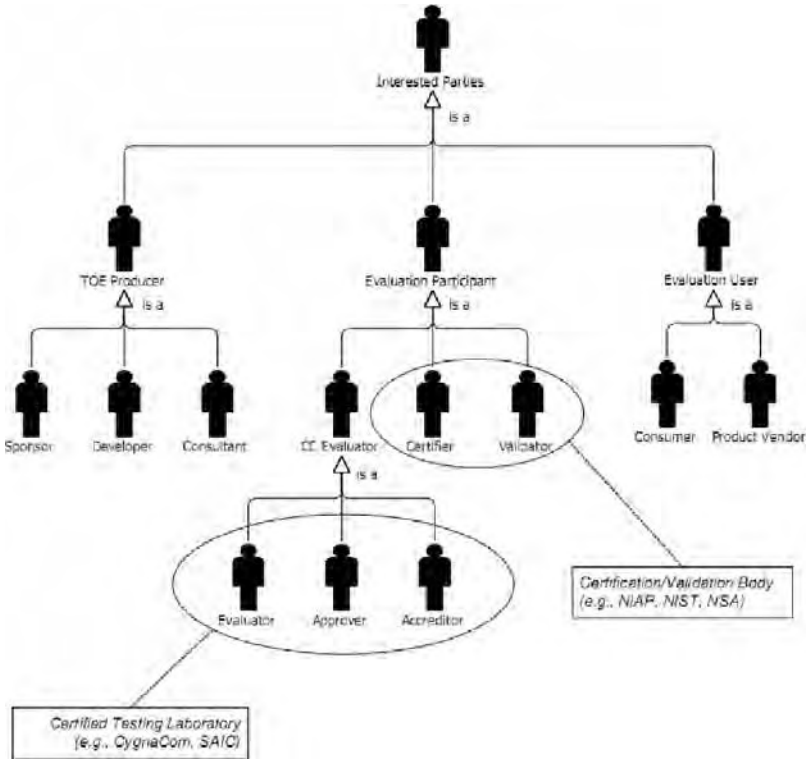


Рис. 8. Онтология субъектов, вовлеченных в процесс оценивания гарантий

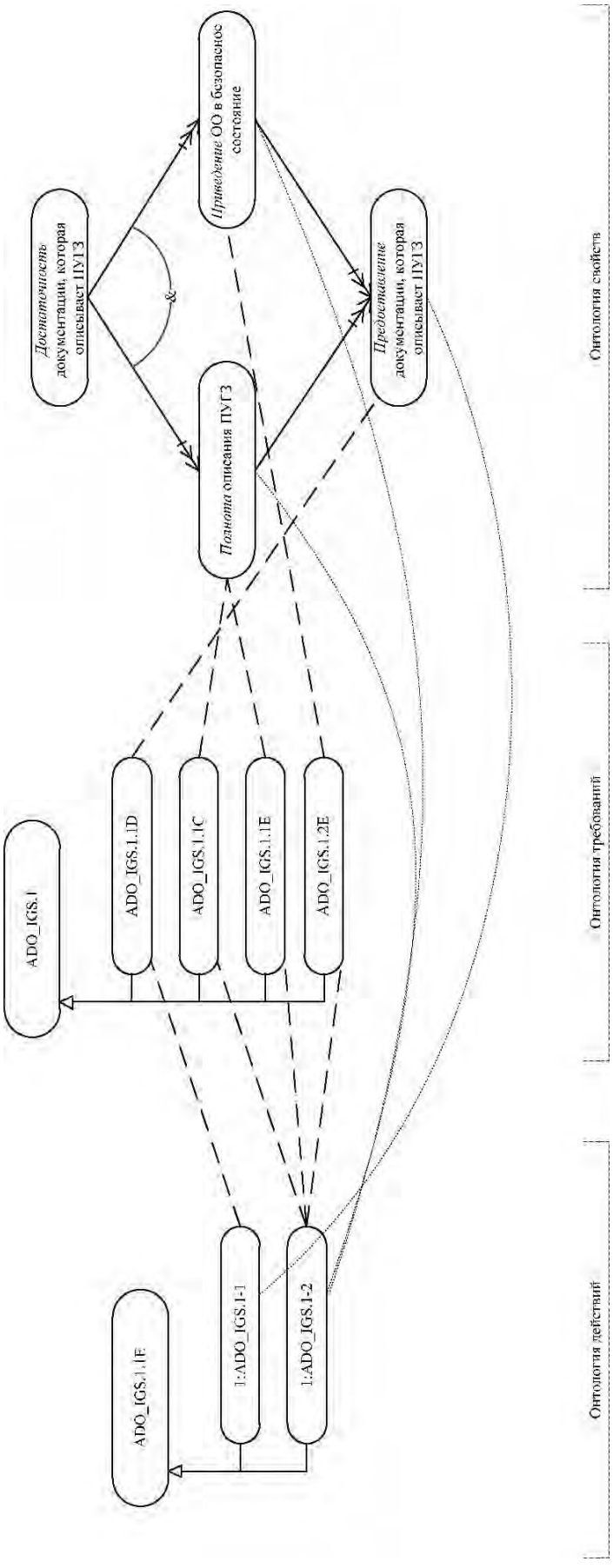


Рис. 9. Процессно-ориентированная онтология по оцениванию документации, которая описывает ПУГЗ

нивания и позволяет конкретизировать основные ее понятия и концепты. Построенные онтологии отображают различные типы связей и зависимостей между концептами области (между требованиями гарантий). Построение онтологических моделей требований гарантий направлено на выполнение требований ширины и глубины оценивания гарантий. Результаты онтологического анализа служат основой для разработки программы оценивания требований гарантий.

Применение функционально-лингвистического подхода для оценивания уровня гарантий безопасности позволяет удовлетворить требования как к процессу оценивания (ширина, глубина и строгость), так и к результатам оценивания (объективность, повторяемость, сопоставимость).

Актуальным вопросом остается разработка инструментальных средств поддержки работы эксперта по проведению оценивания гарантий. Предложенный подход служит основой для проектирования таких инструментальных средств. Дальнейшие исследования могут быть направлены на углубление и уточнение этапов подхода, его развитие и практическую реализацию.

СПИСОК ЛИТЕРАТУРЫ

1. ISO/IEC 15408-1:2005, Informational technology – Security techniques – Evaluation criteria for IT security – Part 1: Introduction and general model.
2. ISO/IEC 15408-2:2005, Informational technology – Security techniques – Evaluation criteria for IT security – Part 2: Security functional requirements.
3. ISO/IEC 15408-3:2005, Informational technology – Security techniques – Evaluation criteria for IT security – Part 3: Security assurance requirement.
4. ISO/IEC 18045:2005, Informational technology – Security techniques – Methodology for IT security evaluation.
5. Критерії оцінки захищеності інформації в комп'ютерних системах від несанкціонованого доступу : НД ТЗІ 2.5-004-99. – [Чинний від 1999-04-22]. – К. : ДСТСЗІ СБ України, 1999. – 53 с. – (Нормативний документ системи технічного захисту інформації).
6. Методичні вказівки з оцінювання функціональних послуг безпеки в засобах захисту інформації від несанкціонованого доступу : НД ТЗІ 2.7-009-09. – [Чинний від 2009-07-24]. – К. : Адміністрація держспецзв'язку, 2009. – 171 с. – (Нормативний документ системи технічного захисту інформації).
7. Методичні вказівки з оцінювання рівня гарантій коректності реалізації функціональних послуг безпеки в засобах захисту інформації від несанкціонованого доступу : НД ТЗІ 2.7-010-09. – [Чинний від 2009-07-24]. – К. : Адміністрація держспецзв'язку, 2009. – 131 с. – (Нормативний документ системи технічного захисту інформації).
8. *Потий, А. В.* IDEF модели процесса оценки уровня гарантий безопасности / А. В. Потий, Д. С. Комин // Труды Научно-технической конференции с международным участием «Компьютерное моделирование в наукоемких технологіях» (КМНТ-2010), Харьков, 18–21 мая 2010 г. – Х. : ХНУ, 2010 – С. 284–287.
9. *Потий, А. В.* Функционально-лингвистический подход к оценке гарантий информационной безопасности / А. В. Потий, Д. С. Комин // Тезисы докладов XIII Международной научно-практической конференции «Безопасность информации в информационно-телекоммуникационных системах», Киев, 18–21 мая 2010 г. – К. : ГСССЗЩ, 2010 – С. 88.
10. *Потій, А. В.* Застосування функціонально-лінгвістичного підходу для оцінювання гарантій інформаційної безпеки / А. В. Потій, Д. С. Комін // Спеціальні телекомунікаційні системи та захист інформації. – 2010. – № 1(17). – С. 24–31.
11. *Потий, А. В.* Нечеткий логический вывод в задачах оценки уровня гарантий безопасности / А. В. Потий, Д. С. Комин // Тези доповідей V Міжнародної науково-практичної конференції «Сучасні проблеми і досягнення в галузі радіотехніки, телекомунікацій та інформаційних технологій», Запоріжжя, 22–24 вересня 2010 р. – Запоріжжя : ЗНТУ, 2010 – С. 121–123.
12. Information Integration for Concurrent Engineering (IICE). IDEF5 Method Report [Электронний ресурс] / Knowledge Based Systems, Inc. – Электрон. дан. – Texas : Knowledge Based Systems , 1994. – Режим доступу: http://www.idef.com/pdf/Idef5.pdf, вільний. – Англ.
13. *Prieto-Diaz, R.* The Common Criteria Evaluation Process. Process Explanation, Shortcomings, and Research Opportunities [Електронний ресурс] / Ruben Prieto-Diaz. – Электрон. дан. – Harrisonburg : Commonwealth Information Security Center, 2002. – (Commonwealth Information Security Center Technical Report Series / Series Editor Samiuel T. Redwine, Jr.) – Режим доступу: https://users.cs.jmu.edu/prietorx/Public/CCevaluationProcesTR03-5.pdf, вільний. – Англ.

Надійшла 17.11.2010

Потій О. В., Комін Д. С.

ОНТОЛОГІЧНЕ МОДЕЛЮВАННЯ ПРОЦЕСУ ОЦІНЮВАННЯ ГАРАНТІЙ В КОНТЕКСТІ ФУНКЦІОНАЛЬНО-ЛІНГВІСТИЧНОГО ПІДХОДУ

Пропонується функціонально-лінгвістичний підхід до оцінювання гарантій безпеки. Наводяться методика та результати онтологічного аналізу предметної області оцінювання гарантій. Обґрунтовується актуальність застосування апарату онтологічного моделювання для завдань оцінювання безпеки.

Ключові слова: гарантії, оцінювання, онтологічне моделювання, функціональне моделювання, лінгвістичні змінні.

Potij A. V., Komin D. S.

ONTOLOGICAL MODELING OF ASSURANCE EVALUATION IN THE CONTEXT OF FUNCTIONAL-LINGUISTIC APPROACH

The functional-linguistic approach to security assurance evaluation is proposed. The procedure and results of assurance evaluation ontological analysis are described. Ontological modeling application to security evaluation is justified.

Key words: assurance, evaluation, ontological modeling, functional modeling, linguistic variables.

УДК 519.832.3

Романюк В. В.

Канд. техн. наук, доцент Хмельницького національного університету

РОЗВ'ЯЗУВАННЯ НЕЧІТКОЇ АНТАГОНІСТИЧНОЇ 2×2-ГРИ

Представлено концепцію розв'язування антагоністичної 2×2-гри, елементи матриці якої задаються у формі неодноелементних множин. Показано, що розв'язком такої нечіткої гри може бути спеціальний перетин розв'язків усіх звичайних 2×2-ігор, елементи матриць яких утворюють ці множини. Для випадків, коли такий перетин виявиться порожнім, пропонується використання нечіткого розв'язку нечіткої 2×2-гри. За умови неприйнятності подібного розв'язку будується безкоаліційна метагра, розв'язок якої міститиме оптимальні поведінки обох гравців у вихідній нечіткій 2×2-грі.

Ключові слова: моделювання в умовах невизначеності, прийняття рішень в умовах невизначеності, нечітка 2×2-гра, безкоаліційна метагра, оптимальна поведінка.

ВСТУП

Прийняття рішень і моделювання в умовах невизначеності [1, 2] є звичною справою, якщо ставиться задача описати і дослідити певне явище або процес з достатньою для практики точністю й адекватністю. Там, де доводиться оцінювати параметри досліджуваного об'єкта, котрі не залежать один від одного і задані у формі інтервалів ненульової міри, застосовані принципи інтервального аналізу [3, 4]. Проте у випадках, коли хоча б два параметри досліджуваного об'єкта є взаємозалежними (не у строго функціональному сенсі), треба адаптовувати методи інтервального аналізу до відповідних нестрого функціональних залежностей. Подібні проблеми виникають і в задачах оптимального керування, частинним випадком яких є ігрові антагоністичні моделі [5], параметри яких задаються нечітко (або ж, іншими словами, ці параметри не можуть бути чітко, у формі точкового значення, оцінені).

АНАЛІЗ ОСТАННІХ ДОСЛІДЖЕНЬ І ПУБЛІКАЦІЙ

Питання нечіткого моделювання і прийняття рішень в умовах повної або часткової невизначеності досить ґрунтовно розглядаються в [1, 2, 6, 7]. Утім, розв'язування ігрових антагоністичних моделей, у яких задані на інтервалах параметри не мають відповідних імовірнісних мір, є невирішеною задачею. Одним зі способів локалізації або точкового оцінювання таких інтервальних параметрів є застосування розв'язку відповідної антагоністичної гри, ядро якої задається на квадраті можливих пар значень параметра, де мінімізуються втрати від некоректної точкової оцінки [8, 9]. Та цей спосіб є занадто песимістичним. Більш того, при розв'язуванні елементарних антагоністичних ігор, тобто 2×2-ігор, де принаймні один з чотирьох елементів матриці гри заданий у формі інтервалу (або сегмента), варто враховувати, що розв'язок матричної гри нетривіальним чином залежить від інтервальних елементів матриці гри [5, 8, 10]. Все це стимулює до формулювання більш простих або до вироблення менш неявних принципів розв'язування матричних ігор з інтервальними (нечіткими) елементами (нечітких ігор). І слід почати, очевидно, з розв'язування нечітких антагоністичних 2×2-ігор.

ФОРМУЛЮВАННЯ МЕТИ ТА ЗАВДАНЬ СТАТТІ

Розв'язок 2×2-гри позначатимемо як $S = \langle \mathbf{P}_{\text{opt}}, \mathbf{Q}_{\text{opt}} \rangle$ з оптимальними стратегіями першого

$$\mathbf{P}_{\text{opt}} = \left[p_{\text{opt}} \; 1 - p_{\text{opt}} \right] \in \in \{ \mathbf{P} = [p \quad 1-p] \in \mathbb{R}^2 \,|\, p \in [0;1] \} \qquad (1)$$

та другого

$$\mathbf{Q}_{\text{opt}} = \left[q_{\text{opt}} \; 1 - q_{\text{opt}} \right] \in \in \{ \mathbf{Q} = [q \quad 1-q] \in \mathbb{R}^2 \,|\, q \in [0;1] \} \qquad (2)$$

гравців відповідно. Вважатимемо, що порядок слідування елементів множини $S = \langle \mathbf{P}_{\text{opt}}, \mathbf{Q}_{\text{opt}} \rangle$ міняти не можна, а усі операції над такими множинами (кортежами) мають здійснюватись окремо для кожного з двох їх елементів. В результаті проведення операцій над множинами виду $S = \langle \mathbf{P}_{\text{opt}}, \mathbf{Q}_{\text{opt}} \rangle$ з'являтимуться двоелементні кортежі, першим елементом яких буде множина з частин оптимальних стратегій першого гравця, а другим – множина з частин оптимальних стратегій другого гравця. Поставимо за мету формалізувати випадок, коли принаймні один з чотирьох елементів матриці 2×2-гри заданий у формі неодноелементної множини (заданий нечітко, але не на

нечіткій множині). Слід запропонувати концепцію розв'язування таких елементарних антагоністичних ігор, де принаймні один з чотирьох елементів матриці 2×2-гри задається нечітко.

ОСНОВНА ЧАСТИНА

Означення 1. Матричну 2×2-гру з ядром

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \tag{3}$$

назвемо нечіткою, якщо $\exists i \in \{1, 2\}$ та $\exists j \in \{1, 2\}$ такі, що $a_{ij} \in A_{ij}$ при $|A_{ij}| > 1$.

Зауваження 1. Якщо усі чотири множини із $\left\{ \{A_{ij}\}_{i=1}^{2} \right\}_{j=1}^{2}$ є скінченними, то нечітку 2×2-гру можна називати дискретно нечіткою.

Зауваження 2. По аналогії з визначенням цілком змішаної гри (або цілком змішаної стратегії гравця) нечітку 2×2-гру, у якій $|A_{ij}| > 1$ виконується $\forall i = \overline{1, 2}$ та $\forall j = \overline{1, 2}$, називатимемо цілком нечіткою.

Зауваження 3. У випадку, коли у системі $\left\{ \{A_{ij}\}_{i=1}^{2} \right\}_{j=1}^{2}$ знайдеться хоча б одна одноелементна множина, відповідну нечітку 2×2-гру можна називати локально нечіткою. Поняття глобально нечіткої 2×2-гри, зрозуміло, співпадає з поняттям цілком нечіткої 2×2-гри.

Зауваження 4. Тривіальна локально нечітка 2×2-гра є найпростішою (елементарною) нечіткою 2×2-грою, де лише одна множина із $\left\{ \{A_{ij}\}_{i=1}^{2} \right\}_{j=1}^{2}$ містить лише два елементи, а решта множин є одноелементними.

Означення 2. Якщо $A_{ij} = [a_{ij}^{\langle 1 \rangle}; a_{ij}^{\langle 2 \rangle}]$ і різниця $a_{ij}^{\langle 2 \rangle} - a_{ij}^{\langle 1 \rangle}$ є постійною $\forall i = \overline{1, 2}$ та $\forall j = \overline{1, 2}$, то 2×2-гру назвемо нечіткою з радіусом нечіткості $r = \dfrac{a_{ij}^{\langle 2 \rangle} - a_{ij}^{\langle 1 \rangle}}{2}$.

Зауваження 5. У випадку, коли 2×2-гра є дискретно нечіткою й $A_{ij} = \{a_{ij}^{\langle 1 \rangle}, a_{ij}^{\langle 2 \rangle}\}$ при $a_{ij}^{\langle 1 \rangle} - a_{ij}^{\langle 2 \rangle}$ $\forall i = \overline{1, 2}$ та $\forall j = \overline{1, 2}$ з постійною різницею $a_{ij}^{\langle 2 \rangle} - a_{ij}^{\langle 1 \rangle}$, теж можна вести мову про застосування означення 2. Крім того, кожна множина A_{ij}, будучи скінченною, може містити більше двох елементів, але так, що $\min A_{ij} = a_{ij}^{\langle 1 \rangle}$ та $\max A_{ij} = a_{ij}^{\langle 2 \rangle}$. Тоді й тут ніщо не перешкоджатиме оперувати радіусом нечіткості $r = \dfrac{a_{ij}^{\langle 2 \rangle} - a_{ij}^{\langle 1 \rangle}}{2}$, причому навіть там, де потужності множин у системі $\left\{ \{A_{ij}\}_{i=1}^{2} \right\}_{j=1}^{2}$ будуть різними.

Тепер час означити те, що надалі вважатимемо розв'язком нечіткої 2×2-гри.

Означення 3. Якщо $S(a_{11}, a_{12}, a_{21}, a_{22})$ є розв'язком 2×2-гри з матрицею (3), то розв'язком нечіткої 2×2-гри назвемо множину (кортеж)

$$\bar{S}(A_{11}, A_{12}, A_{21}, A_{22}) = \bigcap_{a_{11} \in A_{11}} \bigcap_{a_{12} \in A_{12}} \bigcap_{a_{21} \in A_{21}} \bigcap_{a_{22} \in A_{22}} S(a_{11}, a_{12}, a_{21}, a_{22}) = \langle \bar{P}_{\text{opt}}, \bar{Q}_{\text{opt}} \rangle. \tag{4}$$

Зауваження 6. Зрозуміло, що далеко не кожна нечітка 2×2-гра має непорожній розв'язок. Розглянемо хоча б локально нечітку гру з ядром

$$\begin{pmatrix} a_{11} & 4 \\ 3 & 5 \end{pmatrix}, \tag{5}$$

у якому $a_{11} = \{a_{11}^{\langle 1 \rangle}, a_{11}^{\langle 2 \rangle}\}$ (приклад гри з тривіальною нечіткістю). Тут

$$p_{\text{opt}} = \frac{5 - 3}{a_{11} + 5 - 3 - 4} = \frac{2}{a_{11} - 2}$$

при $a_{11} \geq 4$, при $a_{11} \in (3;4)$ імовірність $p_{\text{opt}} = 1$, при $a_{11} = 3$ імовірність $p_{\text{opt}} \in \{1, 0\}$, а при $a_{11} < 3$ імовірність $p_{\text{opt}} = 0$. Також

$$q_{\text{opt}} = \frac{5 - 4}{a_{11} + 5 - 3 - 4} = \frac{1}{a_{11} - 2}$$

при $a_{11} \geq 4$, а при $a_{11} < 4$ імовірність $p_{\text{opt}} = 1$. І, очевидно, що, скажімо, для $a_{11} \in \{3.5, 4.5\}$ розв'язок нечіткої 2×2-гри з матрицею (5)

$$\begin{aligned} \bigcap_{a_{11} \in \{3.5,\, 4.5\}} S(a_{11}, 4, 3, 5) &= \langle \begin{bmatrix} 1 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \end{bmatrix} \rangle \cap \\ \cap \langle \begin{bmatrix} \frac{2}{4.5-2} & 1 - \frac{2}{4.5-2} \end{bmatrix}, \begin{bmatrix} \frac{1}{4.5-2} & 1 - \frac{1}{4.5-2} \end{bmatrix} \rangle &= \\ = \langle \begin{bmatrix} 1 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \end{bmatrix} \rangle \cap \langle \begin{bmatrix} \frac{4}{5} & \frac{1}{5} \end{bmatrix}, \begin{bmatrix} \frac{2}{5} & \frac{3}{5} \end{bmatrix} \rangle &= \\ = \langle \left\{ \begin{bmatrix} 1 & 0 \end{bmatrix} \right\} \cap \left\{ \begin{bmatrix} \frac{4}{5} & \frac{1}{5} \end{bmatrix} \right\}, \left\{ \begin{bmatrix} 1 & 0 \end{bmatrix} \right\} \cap \left\{ \begin{bmatrix} \frac{2}{5} & \frac{3}{5} \end{bmatrix} \right\} \rangle &= \\ = \langle \varnothing, \varnothing \rangle. \end{aligned} \tag{6}$$

Означивши розв'язок нечіткої 2×2-гри, дамо твердження про необхідну умову того, щоб він був непорожнім.

Теорема 1. Для того, щоб розв'язок (4) нечіткої 2×2-гри був непорожнім, необхідно, щоб кожен розв'язок $S(a_{11}, a_{12}, a_{21}, a_{22})$ був у чистих стратегіях.

Доведення. Нехай матрицею 2×2-гри є (3), де немає сідлових точок у чистих стратегіях, а оптимальними стратегіями першого та другого гравців є (1) і (2) відповідно, причому для визначеності вважатимемо, що оптимальна імовірність

$$p_{\text{opt}} = \frac{a_{22} - a_{21}}{a_{11} + a_{22} - a_{21} - a_{12}}$$

вибору першим гравцем його першої чистої стратегії належить інтервалу $(0;1)$. Власне, це означає, що оптимальна стратегія першого гравця є змішаною (або, що характерне для ігор з двома чистими стратегіями даного гравця, є цілком змішаною). При зміні одного зі значень $\left\{\{a_{ij}\}_{i=1}^{2}\right\}_{j=1}^{2}$ імовірність p_{opt} буде змінюватись також. Виключення становитиме випадок, коли $a_{11} = a_{12}$ і змінюватиметься a_{21} або a_{22}. Але тоді $p_{\text{opt}} = 1$ (вибір першої чистої стратегії першим гравцем) або $p_{\text{opt}} = 0$, а

$$q_{\text{opt}} = \frac{a_{22} - a_{12}}{a_{11} + a_{22} - a_{21} - a_{12}} = \frac{a_{22} - a_{12}}{a_{22} - a_{21}}$$

змінюватиметься при $a_{12} \neq a_{21}$ (без втрати загальності). Якщо ж $a_{12} = a_{21}$, то $q_{\text{opt}} = 1$ і гра міститиме сідлову точку у чистих стратегіях. Отже, якщо 2×2-гра розв'язуватиметься у змішаних стратегіях, то кожен раз буде нова множина із множин оптимальних стратегій гравців. Перетин таких множин, очевидно, буде порожнім. Теорему доведено.

Дамо тепер таке означення, що має відношення до непорожнього розв'язку нечіткої 2×2-гри.

Означення 4. Розв'язок 2×2-гри з матрицею (3) назвемо стійким з радіусом r, якщо відповідна нечітка 2×2-гра з елементами $a_{ij} \in [\tilde{a}_{ij} - r; \tilde{a}_{ij} + r]$ має непорожній розв'язок.

Відповідь на питання про існування розв'язків у чистих стратегіях певного класу нечітких 2×2-ігор дає наступне твердження.

Теорема 2. Для довільної 2×2-гри, елементи $\left\{\{a_{ij}\}_{i=1}^{2}\right\}_{j=1}^{2} = \{a, b, c, d\}$ матриці якої є різними, з розв'язком у чистих стратегіях знайдеться $r > 0$ таке, що відповідна нечітка 2×2-гра розв'язуватиметься у чистих стратегіях. При цьому, якщо для визначеності покладати

$$a < b < c < d, \tag{7}$$

значення $r \in (0; r_{\max})$ при

$$r_{\max} = \min\left\{\frac{b-a}{2}, \frac{c-b}{2}, \frac{d-c}{2}\right\}. \tag{8}$$

Доведення. Нехай $a < b < c < d$ і відповідна 2×2-гра має сідлову точку у чистих стратегіях. Зрозуміло, що при заміні a на a_1, b на b_1, c на c_1, d на d_1 такій, що

$$a_1 < b_1 < c_1 < d_1, \tag{9}$$

сідлова точка не зміниться. Візьмемо деяке $r > 0$ і вимагатимемо, щоб

$$a + r < b - r, \quad b + r < c - r, \quad c + r < d - r.$$

Тоді

$$a_1 = a + r, \quad b_1 \in [b - r; b + r],$$
$$c_1 \in [c - r; c + r], \quad d_1 = d - r$$

і буде виконуватись (9). Маємо

$$r < \frac{b-a}{2}, \tag{10}$$

$$r < \frac{c-b}{2}, \tag{11}$$

$$r < \frac{d-c}{2}. \tag{12}$$

Оскільки усі елементи множини $\{a, b, c, d\}$ є різними і, взагалі, має місце (7), то праві частини у (10)–(12) є додатними, і розв'язком системи нерівностей (10)–(12) разом із умовою $r > 0$ є множина

$$\left(0; \min\left\{\frac{b-a}{2}, \frac{c-b}{2}, \frac{d-c}{2}\right\}\right). \tag{13}$$

Теорему доведено.

Звісно, тоді, коли нечітка 2×2-гра не має розв'язку, слід визначати її розв'язок в іншій формі.

Означення 5. Нечітким розв'язком нечіткої 2×2-гри назвемо множину (кортеж)

$$\tilde{S}(A_{11}, A_{12}, A_{21}, A_{22}) =$$
$$= \bigcup_{a_{11}\in A_{11}} \bigcup_{a_{12}\in A_{12}} \bigcup_{a_{21}\in A_{21}} \bigcup_{a_{22}\in A_{22}} S(a_{11}, a_{12}, a_{21}, a_{22}) =$$
$$= \langle \tilde{P}_{\text{opt}}, \tilde{Q}_{\text{opt}} \rangle. \tag{14}$$

Зауваження 7. До такого типу розв'язку доводитиметься звертатись, якщо $S(a_{11}, a_{12}, a_{21}, a_{22})$ є розв'язком 2×2-гри з матрицею (3) й $\bar{S}(A_{11}, A_{12}, A_{21}, A_{22}) = \langle \varnothing, \varnothing \rangle$. Проте про практичну «сумісність» нечіткого розв'язку (14) може йти мова лише у випадках зв'язності множин оптимальних імовірностей p_{opt} та q_{opt} як елементів кортежу $\tilde{S}(A_{11}, A_{12}, A_{21}, A_{22})$ з умовою того, що їх лебегівська міра на числовій прямій буде набагато меншою

за одиницю. У випадку незв'язності одного з елементів кортежу $\tilde{S}(A_{11}, A_{12}, A_{21}, A_{22})$ використання концепції нечіткого розв'язку нечіткої 2×2-гри представляється скрутним. Повертаючись до прикладу локально нечіткої гри з ядром (5), для якої незв'язність множин у кортежі

$$\tilde{S}(A_{11}, A_{12}, A_{21}, A_{22}) =$$
$$= \langle \left[1\ 0\right], \left[1\ 0\right] \rangle \cup \langle \left[\frac{4}{5}\ \frac{1}{5}\right], \left[\frac{2}{5}\ \frac{3}{5}\right] \rangle =$$
$$= \langle \left\{ \left[1\ 0\right] \right\} \cup \left\{ \left[\frac{4}{5}\ \frac{1}{5}\right] \right\}, \left\{ \left[1\ 0\right] \right\} \cup \left\{ \left[\frac{2}{5}\ \frac{3}{5}\right] \right\} \rangle$$

очевидна, тобто перший гравець має брати $p_{\text{opt}} \in \left\{\frac{4}{5}, 1\right\}$, а другий – має брати $q_{\text{opt}} \in \left\{\frac{2}{5}, 1\right\}$, можна говорити про неприйнятно велику відстань між елементами множини $\left\{\frac{2}{5}, 1\right\}$ (навіть відносно її першого елемента). Для множини $\left\{\frac{4}{5}, 1\right\}$ першого гравця відносна відстань між її елементами є значно меншою, тому нечіткий розв'язок тут є більш прийнятним для першого гравця.

Взагалі кажучи, прийнятність нечіткої множини оптимальних стратегій (у формі оптимальних імовірностей вибору першої чистої стратегії) гравця визначається різницею між її максимальним і мінімальним значеннями. Зокрема, чим менша різниця $\sup \tilde{P}_{\text{opt}} - \inf \tilde{P}_{\text{opt}}$, тим більш прийнятною для першого гравця є нечітка множина $\tilde{P}_{\text{opt}}$. Аналогічно і з різницею $\sup \tilde{Q}_{\text{opt}} - \inf \tilde{Q}_{\text{opt}}$ для другого гравця. Зрозуміло, що для зв'язних елементів кортежу (14) замість таких різниць, строго кажучи, слід використовувати лебегівську міру відповідних множин. Тільки тоді можна вести мову про прийнятність не тільки самої концепції нечітких розв'язків у нечітких 2×2-іграх, а й про використання цієї концепції.

Проте, з іншого боку, використання гравцями елементів нечітких множин $\tilde{P}_{\text{opt}}$ та $\tilde{Q}_{\text{opt}}$ у кортежі (14) породжуватиме свою гру, розв'язавши яку, вже можна буде не турбуватись про ступінь прийнятності використання нечіткого розв'язку $\langle \tilde{P}_{\text{opt}}, \tilde{Q}_{\text{opt}} \rangle$. Чистою стратегією кожного гравця у такій грі, котра, у певному смислі, стане вже метагрою по відношенню до вихідної нечіткої 2×2-гри, буде припущення про те, що матрицею 2×2-гри є елемент множини

$$\left\{ \left\{ \left\{ \left\{ \left\{ \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \right\}_{a_{11} \in A_{11}} \right\}_{a_{12} \in A_{12}} \right\}_{a_{21} \in A_{21}} \right\}_{a_{22} \in A_{22}} \right\} \tag{15}$$

з відповідними розв'язком

$$S(a_{11}, a_{12}, a_{21}, a_{22}) =$$
$$= \langle \mathbf{P}_{\text{opt}}(a_{11}, a_{12}, a_{21}, a_{22}), \mathbf{Q}_{\text{opt}}(a_{11}, a_{12}, a_{21}, a_{22}) \rangle. \tag{16}$$

При цьому, очевидно, перший гравець робитиме свої припущення незалежно від другого і навпаки. Якщо обидва гравці одночасно «подумали» про одну й ту саму матрицю

$$\left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \right\} \in$$
$$\in \left\{ \left\{ \left\{ \left\{ \left\{ \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \right\}_{a_{11} \in A_{11}} \right\}_{a_{12} \in A_{12}} \right\}_{a_{21} \in A_{21}} \right\}_{a_{22} \in A_{22}} \right\}, \tag{17}$$

то це ще означає, що перший гравець отримає виграш

$$v_{\text{opt}}(a, b, c, d) =$$
$$= [\mathbf{P}_{\text{opt}}(a, b, c, d)] \cdot \begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot [\mathbf{Q}_{\text{opt}}(a, b, c, d)]^{\text{T}}, \tag{18}$$

а другий гравець стільки ж програє. В описуваній метагрі з'явиться ще один гравець – «природа» або непередбачувані «випадкові обставини», які зумовлять у момент часу обирання гравцями їх чистих стратегій (про обирання змішаних стратегій у певний момент часу говорити складно) той чи інший елемент множини (15) як матрицю (з фіксованими елементами) 2×2-гри. Наприклад, у локально нечіткій грі з ядром (5), у якому $a_{11} \in \{3.5, 4.5\}$, кожен з двох гравців і «природа» у певний момент часу можуть вибирати тільки одну з двох матриць

$$\left\{ \begin{pmatrix} 3.5 & 4 \\ 3 & 5 \end{pmatrix}, \begin{pmatrix} 4.5 & 4 \\ 3 & 5 \end{pmatrix} \right\}.$$

Тут, як метагра по відношенню до вихідної нечіткої 2×2-гри, породжуватиметься діадична гра [8, 11, 12] трьох осіб.

Взагалі кажучи, породженою метагрою по відношенню до вихідної нечіткої 2×2-гри буде безкоаліційна $(|A_{11}| \cdot |A_{12}| \cdot |A_{21}| \cdot |A_{22}|) \times (|A_{11}| \cdot |A_{12}| \cdot |A_{21}| \cdot |A_{22}|) \times$

$\times(|A_{11}|\cdot|A_{12}|\cdot|A_{21}|\cdot|A_{22}|)$-гра трьох осіб. Якщо через $\mathbf{M}_h$ позначати чисту стратегію h-го гравця як елемент множини (15), $\mathbf{P}_{\text{opt}}(\mathbf{M}_1)$ та $\mathbf{Q}_{\text{opt}}(\mathbf{M}_2)$ – як оптимальні стратегії в $\mathbf{M}_1$-грі та $\mathbf{M}_2$-грі відповідно, то у такій метагрі функція виграшу першого гравця

$$K_1(\mathbf{M}_1, \mathbf{M}_2, \mathbf{M}_3) = \\ = \mathbf{P}_{\text{opt}}(\mathbf{M}_1)\cdot\mathbf{M}_3\cdot[\mathbf{Q}_{\text{opt}}(\mathbf{M}_2)]^{\mathrm{T}}, \qquad (19)$$

функція виграшу другого гравця

$$K_2(\mathbf{M}_1, \mathbf{M}_2, \mathbf{M}_3) = \\ = -\mathbf{P}_{\text{opt}}(\mathbf{M}_1)\cdot\mathbf{M}_3\cdot[\mathbf{Q}_{\text{opt}}(\mathbf{M}_2)]^{\mathrm{T}} = \\ = -K_1(\mathbf{M}_1, \mathbf{M}_2, \mathbf{M}_3), \qquad (20)$$

а функцію виграшу третього гравця («природи») можна покласти тотожною нулю:

$$K_3(\mathbf{M}_1, \mathbf{M}_2, \mathbf{M}_3) = 0,$$

$$\forall\mathbf{M}_h \in \left\{\left\{\left\{\left\{\left\{\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22}\end{pmatrix}\right\}_{a_{11}\in A_{11}}\right\}_{a_{12}\in A_{12}}\right\}_{a_{21}\in A_{21}}\right\}_{a_{22}\in A_{22}}\right\},$$

$$h = \overline{1,3}. \qquad (21)$$

Очевидно, що для скінченних множин A_{11}, A_{12}, A_{21}, A_{22} функції виграшів (19)–(21) можна представити у формі тривимірних

$$(|A_{11}|\cdot|A_{12}|\cdot|A_{21}|\cdot|A_{22}|)\times(|A_{11}|\cdot|A_{12}|\cdot|A_{21}|\cdot|A_{22}|)\times \\ \times(|A_{11}|\cdot|A_{12}|\cdot|A_{21}|\cdot|A_{22}|)\text{-матриць},$$

але, зважаючи на (20) і (21), можна обмежитись однією

$$(|A_{11}|\cdot|A_{12}|\cdot|A_{21}|\cdot|A_{22}|)\times(|A_{11}|\cdot|A_{12}|\cdot|A_{21}|\cdot|A_{22}|)\times \\ \times(|A_{11}|\cdot|A_{12}|\cdot|A_{21}|\cdot|A_{22}|)\text{-матрицею}$$

$\mathbf{K} = [k_{m_1m_2m_3}]_{G\times G\times G}$ з елементами

$$k_{m_1m_2m_3} = \mathbf{P}_{\text{opt}}(\mathbf{M}_1^{\langle m_1\rangle})\cdot\mathbf{M}_3^{\langle m_3\rangle}\cdot[\mathbf{Q}_{\text{opt}}(\mathbf{M}_2^{\langle m_2\rangle})]^{\mathrm{T}}, \\ m_h = \overline{1,G}\ \ \forall h = \overline{1,3}, \qquad (22)$$

де

$$G = |A_{11}|\cdot|A_{12}|\cdot|A_{21}|\cdot|A_{22}|$$

й

$$\left\{\mathbf{M}_h^{\langle m_h\rangle}\right\}_{m_k=1}^{G} = \\ = \left\{\left\{\left\{\left\{\left\{\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22}\end{pmatrix}\right\}_{a_{11}\in A_{11}}\right\}_{a_{12}\in A_{12}}\right\}_{a_{21}\in A_{21}}\right\}_{a_{22}\in A_{22}}.$$

Так само можна обмежитись лише функцією (19) у випадку, коли принаймні одна з множин системи $\left\{\{A_{ij}\}_{i=1}^{2}\right\}_{j=1}^{2}$ є нескінченною.

Отже, нечітка 2×2-гра з множиною (15) усіх можливих «чітких» 2×2-матриць за умови порожнього кортежу (4) і неприйнятності нечіткого розв'язку (14) породжуватиме безкоаліційну $G\times G\times G$-гру (метагру) з ядром (19) при штучному покладанні (21). Розв'язок цієї метагри міститиме оптимальну поведінку обох гравців в умовах ведення вихідної нечіткої 2×2-гри.

ВИСНОВОК ТА ПЕРСПЕКТИВА ПОДАЛЬШОГО ДОСЛІДЖЕННЯ

Звичайно, питання про впроваджуваність розв'язку безкоаліційної $G\times G\times G$-метагри є відкритим і далеко не тривіальним [8, 11, 12]. Можливо, у деяких ситуаціях практичне використання цього розв'язку [13, 14] виявиться настільки незрозумілим (адже мова йтиме як про імовірності вибору чистих стратегій – матриць, так і про відому неоднозначність ситуацій рівноваги у безкоаліційних іграх), що буде варто шукати способи усунення «нечіткості» у 2×2-грі (прямого, «неігрового» перетворення нечіткої 2×2-гри у «чітку»). Але запропонована концепція розв'язку (4) нечіткої 2×2-гри та її нечіткого розв'язку (14) з можливим розв'язуванням безкоаліційної $G\times G\times G$-метагри має фундаментальне значення для ігрових методів моделювання в умовах невизначеності. У перспективі, звісно, слід зайнятися формулюванням відповідних положень для матричних ігор більших і, взагалі кажучи, довільних форматів.

СПИСОК ЛІТЕРАТУРИ

1. *Трухаев, Р. И.* Модели принятия решений в условиях неопределенности / Трухаев Р. И. – М. : Наука, 1981. – 258 с.
2. *Черноруцкий, И. Г.* Методы принятия решений / Черноруцкий И. Г. – СПб. : БХВ-Петербург, 2005. – 416 с. : ил.
3. *Caprani, O.* Introduction to Interval Analysis / O. Caprani, K. Madsen, H. B. Nielsen. – IMM, DTU, 2002. – 82 p.
4. *Большаков, А. А.* Методы обработки многомерных данных и временных рядов : [учебное пособие для вузов] / А. А. Большаков, Р. Н. Каримов. – М. : Горячая линия – Телеком, 2007. – 520 с. : ил.
5. *Петросян, Л. А.* Теория игр: [учеб. пособие для унтов] / Петросян Л. А., Зенкевич Н. А., Семина Е. А. – М. : Высшая школа ; Книжный дом «Университет», 1998. – 304 с. : ил.
6. *Романюк, В. В.* Мінімаксний підхід у реалізації стохастичного параметра з невідомим імовірнісним розподілом на інтервалі ненульової міри / В. В. Романюк // Вісник Хмельницького національного університету. Технічні науки. – 2010. – № 3. – С. 65–71.

7. *Романюк, В. В.* Оцінювання вірогідності розподілу статистичних частот випадкової величини з невідомим математичним сподіванням і дисперсією / В. В. Романюк // Вісник НТУ «ХПІ». Тематичний випуск: Інформатика і моделювання. – Харків : НТУ «ХПІ», 2010. – № 21. – С. 152–161.
8. *Воробьёв, Н. Н.* Теория игр для экономистов-кибернетиков / Воробьёв Н. Н. – М. : Наука, Главная редакция физико-математической литературы, 1985. – 272 с.
9. *Романюк, В. В.* Модель визначення оптимального рішення проектувальника у задачі про розрахунок повздовжньої стійкості двох елементів будівельної конструкції при дії на них нормованого стискаючого зусилля / В. В. Романюк // Проблеми трибології. – 2010. – № 1. – С. 42–56.
10. *Оуэн, Г.* Теория игр : [пер. с англ.] / Оуэн Г. – 2-е изд. – М. : Едиториал УРСС, 2004. – 216 с.
11. *Романюк, В. В.* Рекомендації щодо використання нерівноважної симетричної ситуації у діадичній грі як моделі охорони навколишнього середовища з трьома суб'єктами забруднення довкілля / В. В. Романюк // Екологічна безпека та природокористування. – 2010. – Вип. 5. – С. 144–159.
12. *Романюк, В. В.* Практична реалізація стратегії у найвигіднішій симетричній ситуації у діадичній грі з трьома суб'єктами забруднення водойми / В. В. Романюк // Екологічна безпека. – 2009. – № 4 (8). – С. 49–56.
13. *Романюк, В. В.* Метод реалізації принципу оптимальності у матричних іграх без сідлової точки / В. В. Романюк // Вісник НТУ «ХПІ». Тематичний випуск: Інформатика та моделювання. – Харків : НТУ «ХПІ», 2008. – № 49. – С. 146–154.
14. *Романюк, В. В.* Метод реалізації оптимальних змішаних стратегій у матричній грі з порожньою множиною сідлових точок у чистих стратегіях з відомою кількістю партій гри / В. В. Романюк // Наукові вісті НТУУ «КПІ». – 2009. – № 2. – С. 45–52.

Надійшла 19.10.2010

Романюк В. В.

РЕШЕНИЕ НЕЧЕТКОЙ АНТАГОНИСТИЧЕСКОЙ 2×2-ИГРЫ

Представлена концепция решения антагонистической 2×2-игры, элементы матрицы которой задаются в форме неодноэлементных множеств. Показано, что решением такой нечеткой игры может быть специальное пересечение решений всех обычных 2×2-игр, элементы матриц которых образуют эти множества. Для случаев, когда такое пересечение окажется пустым, предлагается использование нечеткого решения нечеткой 2×2-игры. При условии неприемлемости подобного решения строится бескоалиционная метаигра, решение которой будет содержать оптимальные поведения обоих игроков в исходной нечеткой 2×2-игре.

Ключевые слова: моделирование в условиях неопределенности, принятие решений в условиях неопределенности, нечеткая 2×2-игра, бескоалиционная метаигра, оптимальное поведение.

Romanuke V. V.

SOLVING THE FUZZY ANTAGONISTIC 2×2-GAME

There has been represented a concept of solving the antagonistic 2×2-game, whose matrix elements are defined in the form of non-one-element sets. It has been revealed that the solution of such fuzzy game may be a special intersection of solutions of all ordinary 2×2-games, whose matrices elements constitute those sets. For cases when such intersection appears to be empty, it is suggested to use a fuzzy solution of a fuzzy 2×2-game. If this solution is unacceptable, a noncooperative metagame is constructed, the solution of which will contain the optimal behavior of both players in the initial fuzzy 2×2-game.

Key words: modeling within uncertainty, decision making within uncertainty, fuzzy 2×2-game, noncooperative metagame, optimal behavior.

УДК 519.816+519.712.6

Федюкович В. Е.

Инженер ООО «ИнтроПро» (г. Киев)

О НЕОБХОДИМОСТИ ДОПОЛНИТЕЛЬНОЙ ПРОВЕРКИ СЕРТИФИКАТА СХЕМЫ DAA

Выполнен анализ схемы DAA. Обнаружено, что схема допускает совместные действия Эмитента и Проверяющего с целью получить дополнительную информацию о Пользователе, которые не обнаруживаются Пользователем, следующим протоколу. Предложена дополнительная проверка Пользователем DAA сертификата, полученного от Эмитента, позволяющая обнаружить такую атаку и прекратить протокол.

Ключевые слова: DAA, анонимность, аутентификация, протокол доказательства знания, TPM.

ВВЕДЕНИЕ

Аппаратное обеспечение персонального компьютера состоит, с момента его появления, из унифицированных блоков, что допускает его самостоятельную сборку. Обратная сторона максимально упрощенной процедуры сборки заключается в фактическом отсутствии механизмов контроля целостности аппаратного обеспечения, что затрудняет обнаружение вмешательства на аппаратном уровне. Такое вмешательство, в свою очередь, может приводить к утечке конфиденциальных данных при их обработке на таком компьютере. Контроль целостности аппаратного и программного обеспечения компьютера является одной из основных целей, решаемых в рамках Trusted Computing

Group [1] (TCG). TCG анонсировала цели сохранности важной информации путем создания защищенного носителя данных; создания механизмов надежной аутентификации компьютеров, в том числе удаленной; контроля целостности и управления пользователем путей предоставления информации о целостности третьим лицам. Задача контроля целостности решается путем «измерения» параметров старта компьютера и сохранения их в микросхеме Trusted Platform Module (TPM). Предусмотрен механизм «накопления»: в регистр TPM помещается значение хэш-функции, аргументами которой являются текущее содержание регистра и очередное измерение. TPM предоставляет доступ на чтение к значению регистра, содержащему произвольные данные пользователя, при условии совпадения состояния такого компьютера, представленного регистром накопления, с состоянием на момент инициализации такого регистра. TCG также имеет надежную проверку состояния удаленного компьютера, с учетом ожидаемого конфликта интересов сторон, связанного с распространением персональных данных владельца компьютера. В версии 1.1 спецификаций TPM используется цифровая электронная подпись, выполняемая потенциально короткоживущими RSA ключами Attestation Identity Key (AIK). Такие ключи заверяются доверенной третьей стороной (Privacy CA), которая, в свою очередь, аутентифицирует компьютер по постоянному RSA ключу Endorsement Key (EK). Версия 1.2 спецификаций предусматривает схему Direct Anonymous Attestation (DAA) [2], в которой для скрытия связи между экземпляром подписи доверенной стороны и сертификатом пользователя используется вариант механизма затемнения (blinding) [3], полностью выполняемый на уровне программного обеспечения компьютера пользователя. Имеется свойство анонимности схемы: различные экземпляры подписи, созданные Пользователем, а также различными Пользователями, неотличимы.

Микросхема TPM серийно выпускается Infineon и другими компаниями, и устанавливается на некоторые системные платы и ноутбуки. Функциональность TPM включена в микросхему южного моста некоторых наборов логики (chipset) Intel. Дальнейшее развитие [4] схемы DAA предполагает использование эллиптических кривых, имеющих билинейные отображения (bilinear pairing).

Инициатива Trusted Computing в целом подверглась критике [5, 6] со стороны Фонда свободного программного обеспечения (FSF). Следует отметить, что ряд утверждений можно рассматривать как предположения о намерениях и планах участниках рынка. Так, например, в эссе Столлмена [5] содержится предположение о рисках для пользователей компьютеров, связанных с потерей возможности устанавливать и использовать свободное программное обеспечение; при этом делается ссылка на законодательные инициативы в США. Отдельного внимания заслуживает ссылка на программное обеспечение GNU Privacy Guard (GPG), а также утверждение о полезности GPG при пересылке информации по электронной почте, в форме противопоставления функциональности GPG и предполагаемых целей Trusted Computing. В материале Андерсена [6] корректно изложена идея мониторинга старта компьютера, на основании чего делается ряд предположений, в том числе о возможности избирательного блокирования компьютера, идентифицируемого на основании уникальных ключей. В статье [7] сформулировано утверждение о рисках, связанных с предоставлением третьим лицам точной информации о программном обеспечении Пользователя.

Ожидание анонимности пользователей при удаленной проверке целостности их компьютеров в рамках схемы DAA является решающим фактором, объясняющим интерес к изучению такого механизма проверки на уровне серийно выпускаемого оборудования. Ряд утверждений о возможностях TPM и рисках, возникающих при его использовании, следует рассматривать как необоснованные, а также игнорирующие возможности, предоставляемые протоколами доказательства знания для ограничения распространения персональной информации. С другой стороны, следует обратить внимание на ошибку, нередко встречающаяся при проектировании программного обеспечения: недостаточная проверка возвращаемого значения, что особенно важно в случае совместных вычислений и конфликта интересов участников вычислений. В этой работе изложены результаты независимого анализа схемы DAA, которые могут быть полезны при анализе рисков и выработке рекомендаций.

1. ОБЩАЯ ИНФОРМАЦИЯ О СХЕМЕ DAA

Участниками схемы являются Пользователи, Проверяющие и Эмитент. Схема состоит из алгоритма выбора параметров схемы (Setup), протокола выпуска Эмитентом сертификата Пользователя (Join), алгоритмов создания и проверки экземпляра подписи Пользователя (Sign и Verify). Пользователь создает экземпляр подписи, который является неинтерактивным вариантом [8] протокола доказательства знания ключей TPM, таких, что имеется экземпляр подписи Эмитента на экземпляре привязки к этим ключам. Предусматривается генерация микросхемой случайных значений (nonce), которые являются дополнительными аргументами хэш-функции при выборе значения запроса, отсутствующими в оригинальной работе [8]. Все вычисления с ключами TPM выпол-

няются на уровне микросхемы (TPM), операции с затемнением (blinding) и обмен cmoe e и случайное v'', вычисляет элемент группы A, такой, что

$$A^e US^{v''} = Z \pmod{n}, \quad (1)$$

где S, Z – элементы группы, параметры схемы. Используется подгруппа квадратичных вычетов мультипликативной группы кольца вычетов по модулю n для составного n, выбранного Эмитентом на этапе генерации параметров схемы. Эмитент пересылает пользователю (A, e, v''), что является экземпляром подписи вида Camenisch-Lysyanskaya [9]. Для создания экземпляра такой подписи необходимо знание факторизации n, что является ключом Эмитента. Исчерпывающая информация о схеме DAA приведена в оригинальной работе [1].

2. РЕЗУЛЬТАТЫ

Было замечено, что Пользователь использует полученные от Эмитента данные без предварительной проверки, удовлетворяют ли они уравнению (1) как экземпляр подписи Эмитента. Был обнаружен сценарий для Эмитента и Проверяющего (в дальнейшем называемых Соперником), в рамках которого Пользователь не достигает неотличимости событий аутентификации. Пусть Эмитент создал экземпляр подписи, который удовлетворяет уравнению (1) для некоторого произвольного уникального $\bar{Z}$, отличного от Z. Тогда Пользователь, следующий спецификациям, создает экземпляр подписи, успешно проверяемый уравнением (в обозначениях оригинальной работы)

$$T_1^e R_0^{f_0} R_1^{f_1} S^v h^{-ew} = \bar{Z} \pmod{n}, \quad (2)$$

Проверяющий, следующий спецификациям, отвергает такой экземпляр подписи как некорректный, так как значение запроса не совпадает со значением хэш-функции. Однако Проверяющий, имеющий список значений $\bar{Z}$, полученный от Эмитента, может попытаться воссоздать значение запроса, перебирая все значения из такого списка. А именно, при таком переборе следует использовать $\bar{T}_1$ при формировании аргумента хеш-функции вместо $\hat{T}_1$ в оригинальной работе:

$$\bar{T}_1 = \bar{Z}^{-c} T_1^{s_e + c2^{l_e-1}} R_0^{s_{f_0}} R_1^{s_{f_1}} S^{s_v} h^{-s_{ew}} \pmod{n}. \quad (3)$$

Таким образом, Проверяющий всегда может распознать сертификат, выданный Пользователю действующим произвольно Эмитентом, при условии выдачи уникальных некорректных (т. е. не удовлетворяющих уравнению (1) проверки подписи) сертификатов. Такая конструкция позволяет такому Сопернику формировать историю событий аутентификации выбранных Эмитентом Пользователей, что делает заявленное свойство анонимности схемы DAA требующим уточнения. Схема DAA также допускает Пользователя, который всегда выполняет дополнительную проверку (1) полученного от Эмитента сертификата. Такой Пользователь всегда обнаруживает попытку нарушения анонимности путем предоставления некорректного сертификата и может прекратить протокол с таким Эмитентом. Кроме того, действия такого Пользователя неотличимы от действий Пользователя, следующего протоколу в случае, если Эмитент также следует протоколу. Дополнительная проверка предусматривает реализацию на уровне программного обеспечения и не требует каких-либо изменений в микросхеме Trusted Platform Module (TPM). Эти результаты были изложены в препринте IACR [10] и представлены на конференции РусКрипто [11]. В последовавших работах (например, [12]) предусмотрена проверка Пользователем корректности полученного экземпляра подписи Эмитента.

Необходимо также обратить внимание на дополнительные случайные значения, выбираемые TPM при формировании запроса при помощи хэш-функции. Как следствие, эти случайные значения необходимы для проверки корректности подписи Пользователя, что допускает скрытый канал передачи данных TPM – Проверяющий. Следует отметить, что предложенный способ формирования запроса не допускает механизма совместного выбора случайных значений, предложенного в модели «наблюдатель в кошельке» [13].

Формирование ответов Пользователя в виде экземпляра подписи может ограничивать возможности такого Пользователя в управлении доступностью информации о состоянии компьютера для третьих лиц. Определенный интерес может представлять интерактивная проверка состояния, в том числе с учетом схем с выбранным заранее Проверяющим (designated Verifier) или схем с подтверждением (confirmer signatures).

ВЫВОДЫ

Обнаружена уязвимость схемы Direct Anonymous Attestation в модели угроз, предусматривающей произвольные совместные действия Эмитента и Проверяющего. Уязвимость позволяет такому Сопернику исключить анонимность Пользователя, следующего спецификациям. Выпуск некорректных сертификатов может оставаться незамеченным в случаях использования программного обеспечения Эмитента и Проверяющего только одного производителя. Такая уязвимость может быть исключена дополнительной проверкой Пользователем полученного сертификата. Также отмечена возможность скрытого канала передачи путем генерации микросхемой TPM псев-

дослучайных чисел, которые должны быть получены Проверяющим в неизменном виде.

СПИСОК ЛИТЕРАТУРЫ

1. *Brickell, E.* Direct Anonymous Attestation[Электронный ресурс] / Brickell E., Camenisch J. and Chen L. // Cryptology ePrint Archive. – Report 2004/205. – Режим доступа: http://eprint.iacr.org/2004/205/.
2. Trusted Computing Group [Электронный ресурс]. – Режим доступа: http://www.trustedcomputinggroup.org/.
3. *Chaum, D.* Blind Signatures for Untraceable Payments / Chaum D., Rivest R. L. and Sherman A. T. (Eds.) // Advances in Cryptology : proceedings of CRYPTO'82. – Plemum, New York, 1983. – P. 89–105.
4. *Brickell, E.* Simplified security notions of direct anonymous attestation and a concrete scheme from pairings / Brickell E., Chen L. and Li J. // International Journal of Information Security. – 2009. – Vol. 8. – P. 315–330.
5. *Stallman, R.* Can You Trust Your Computer? [Электронный ресурс] / Richard Stallman // Free Software Free Society: selected essays of Richard M. Stallman. – Режим доступа: http://www.gnu.org/philosophy/can-you-trust.html.
6. *Anderson, R.* 'Trusted Computing' Frequently Asked Questions [Электронный ресурс] / Anderson R. – Режим доступа: http://www.cl.cam.ac.uk/~rja14/tcpa-faq.html.
7. Trusted Computing: Promise and Risk [Электронный ресурс] // Electronic Frontier Foundation whitepaper. – Режим доступа: http://www.eff.org/wp/trusted-computing-promise-and-risk.
8. *Fiat, A.* How to Prove Yourself: Practical Solutions to Identification and Signature Problems / Fiat A. and Shamir A. // Lecture Notes in Computer Science. – 1987. – Vol. 263. – P.186–194.
9. *Camenisch, J.* A Signature Scheme with Efficient Protocols / Camenisch J. and Lysyanskaya A. // Lecture Notes in Computer Science. – 2003. – Vol. 2576. – P.268–289.
10. *Fedyukovych, V.* A strategy for any DAA Issuer and an additional verification by a Host [Электронный ресурс] / V. Fedyukovych // Cryptology ePrint Archive. – Report 2008/277. – Режим доступа: http://eprint.iacr.org/2008/277/.
11. *Федюкович, Е.* Восстановление анонимности при использовании протоколов DAA [Электронный ресурс] / В. Е. Федюкович // Рускрипто 2009. – Режим доступа: http://ruscrypto.ru/sources/conference/rc2009/.
12. *Brickell, E.* Enhanced Privacy ID from Bilinear Pairing [Электронный ресурс] / Brickell E. and Li J. // Cryptology ePrint Archive. – Report 2009/095. – Режим доступа: http://eprint.iacr.org/2009/095/.
13. *Chaum, D.* Wallet Databases with Observers / Chaum D. and Pedersen T. P. // Lecture Notes in Computer Science. – 1993. – Vol. 740/1993. – P. 89–105.

Надійшла 1.11.2010

Федюкович В. Є.

ПРО ДОДАТКОВУ ПЕРЕВІРКУ СЕРТІФІКАТА СХЕМИ DAA

Було виконано аналіз схеми DAA. Було знайдено, що схема не є анонімною: Емітент може випустити сертифікат, який завжди може впізнати Перевіряючий. Також було запропоновано додаткове рівняння перевірки, щоб уникнути такої атаки.

Ключові слова: DAA, анонімність, атрибуція, протокол доказу знання, TPM.

Fedyukovych V.

ON ADDITIONAL VERIFICATION OF DDA CERTIFICATE

A strategy for colluding Issuer and Verifier with DAA scheme was found to let such an adversary always distinguish honest Users that were issued 'tagged' certificates voiding anonymity property of DAA. Additional verification equation was introduced to detect such an attack.

Key words: DAA, anonimity, authentication, proof of knowledge, TPM.

УДК 681.3.06

Халимов Г. З.

Канд. техн. наук, доцент Харьковского национального университета радиоэлектроники

ОЦЕНКА ПАРАМЕТРОВ КРИВЫХ ФЕРМА ДЛЯ УНИВЕРСАЛЬНОГО ХЕШИРОВАНИЯ

Получены точные решения для числа точек кривых Ферма, когда порядок поля имеет делители 2, 3 и 6, также оценки числа точек на основе вероятностного подхода. Приводятся асимптотические границы отношения максимального числа точек кривой Ферма в простом поле к ее роду и к границе Хассе – Вейля.

Ключевые слова: универсальное хеширование, кривые Ферма.

ОБЩАЯ ПОСТАНОВКА ЗАДАЧИ И ЕЕ АКТУАЛЬНОСТЬ

Универсальное хеширование на основе алгеброгеометрического подхода впервые было предложено Биербрауэром и Кабатинским [1]. Пусть задана абсолютно неразложимая, несингулярная проективная кривая χ над полем F_q с точками $P = \{P_1, P_2, \ldots, P_n\} \in \chi(F_q)$. Для каждой алгебраической кривой можно определить поле рациональных функций $F_q(\chi)$. В каждой точке P_j кривой χ можно вычислить оценку ϑ_P для рациональных функций $f_i \in F_q(\chi)$, которая определяет порядок нуля или полюса функции f_i в этой точке. Хеш-значение $h_{P_j}(m) \in F_q$ для сообщения $m = (m_1, m_2, \ldots, m_k)$, $m_i \in F_q$ в точке $P_j \in F_q$ определяется выражением

$$h_{P_j}(m) = \sum_{i=1}^{k} f_i(P_j) m_i, \qquad (1)$$

где $f_i \in F_q(\chi)$ с упорядоченными порядками полюсов $0 < u_1 < u_2 < \ldots < u_k$. Хеш-функция $h_{P_j}(m)$ опреде-

ляет универсальный хеш-класс $\varepsilon - U(N, q^k, q)$, где вероятность коллизии $\varepsilon \le u_k/N$, N – число точек алгебраической кривой, q^k – объем пространства сообщений, q – объем пространства хеш-кодов. Хеш-функция $h_{P_j}(m)$ является ключевой, ее значение зависит от точки P_j кривой χ. Оценка для вероятности коллизии зависит от алгеброгеометрических параметров кривой.

Утверждение 1. [2] Вероятность коллизии при универсальном хешировании (1) при $k > 2g - 1$, где g – род алгебраической кривой и N – число точек кривой, определяется границей

$$\varepsilon \le (k + g - 1)/N. \qquad (2)$$

Проблематика построения схем универсального хеширования на основе алгеброгеометрического представления заключается в выборе алгебраических кривых с требуемыми параметрами, прежде всего с как можно большим отношением числа точек кривой к ее роду, а также с реализацией вычислений в конечном поле F_q, простотой вычислений и согласованностью с представлением информационных данных. Интерес представляют конструкции простых полей с модулями $2^m \pm 1$ или близких к ним простых чисел. В представленных материалах отображены основные результаты исследований по кривым Ферма в простом поле для целей универсального хеширования.

Целью статьи является нахождение оценок для числа решений кривой Ферма в простом поле. В разделе 1 рассмотрены основные свойства кривых Ферма и точные значения числа точек для кривых Ферма в простом поле. В разделе 2 на основе вероятностного подхода получены оценки числа решений уравнения Ферма в конечном поле. В разделе 3 приводятся асимптотические результаты по кривым Ферма над простым полем.

1. ТОЧНЫЕ ЗНАЧЕНИЯ ЧИСЛА ТОЧЕК ДЛЯ КРИВЫХ ФЕРМА В ПРОСТОМ ПОЛЕ

Кривые Ферма Fr_m определяются выражением

$$X^m + Y^m + Z^m = 0, \qquad (3)$$

имеют частные производные вида $F_X = mX^{m-1}$, $F_Y = mY^{m-1}$, $F_Z = mZ^{m-1}$. Рассмотрим основные свойства кривых Ферма для случая простого поля F_q.

Утверждение 1. Пусть кривая Ферма Fr_m определена над простым полем F_q. Справедливо следующее:

1) кривая Fr_m является неприводимой, несингулярной кривой степени m без особенностей, рода $g = \dfrac{(m-1)(m-2)}{2}$;

2) если m взаимно просто с $q - 1$, тогда $X^m + Y^m + Z^m = 0$ изоморфна $X + Y + Z = 0$ и имеет число точек $N = q + 1$;

3) кривая вида $X^{(q-1)/3} + Y^{(q-1)/3} + Z^{(q-1)/3} = 0$ имеет $N = 2(q-1)^2/9$ и $g = (q-4)(q-7)/18$;

4) кривая вида $X^{(q-1)/2} + Y^{(q-1)/2} + Z^{(q-1)/2} = 0$ имеет $N = 3(q-1)/2$ и $g = (q-3)(q-5)/8$;

5) кривая вида $X^{(q-1)/6} + Y^{(q-1)/6} + Z^{(q-1)/6} = 0$ имеет число точек $N = (q-1)/2 + (q-1)^2/18$ и род $g = (q-7)(q-13)/72$;

6) кривая $X^{(q-1)/2^m} + Y^{(q-1)/2^m} + Z^{(q-1)/2^m} = 0$ имеет род $g = (q - 2^m - 1)(q - 2^m - 2)/2^{2m+1}$ и число точек $N = 3(q-1)/2^m$, если $\left|2^{2^{m-1}}\right| \ne 1 (\mathrm{mod} q)$ и $N = 3(q-1)/2^m + 3(q-1)^2/2^{2m}$, если $2^{2^{m-1}} = -1 (\mathrm{mod} q)$.

Результаты 1 и 2 являются очевидными и известными. Результаты 3–6 получаются методом подсчета числа решений для уравнений Ферма на основе свойства суммы элементов мультипликативной подгруппы второго, третьего и шестого порядка. Значение рода кривой определяется формулой Римана – Роха.

Результат 3 следует из того, что решениями уравнения кривой в проективном пространстве P^2 над F_q являются точки $(\gamma : \xi : 1)$, для которых справедливо $\gamma^{(q-1)/3} + \xi^{(q-1)/3} + 1 = 0$. Так как $\gamma, \xi \in F_q$ и $\gamma = \alpha^i$, $\xi = \alpha^j$, α – образующий элемент поля, имеем $\alpha^{i(q-1)/3} + \alpha^{j(q-1)/3} + 1 = 0$ или $\beta^i + \beta^j + 1 = 0$, где $\beta = \alpha^{(q-1)/3}$ – образующий элемент мультипликативной подгруппы третьего порядка. В силу свойства $\sum_{k=0}^{n-1} \beta^k = 0$ для элементов мультипликативной группы порядка n получим, что уравнение Ферма имеет решение, если и только если $\beta^{(i)} = \beta^1$ и $\beta^{(j)} = \beta^2$, где $(\cdot)$ обозначает вычисление значения степени по модулю порядка β, а так же если $\beta^{(i)} = \beta^2$ и $\beta^{(j)} = \beta^1$. Число решений по каждому набору условий равно $(q-1)^2/9$ и $N = 2(q-1)^2/9$.

Результат 4 является очевидным. Элементы поля $\alpha^{i(q-1)/2}$ принадлежат мультипликативной подгруппе второго порядка: $1, \beta = \alpha^{(q-1)/2}$ и $1 + \beta = 0$. Кривая Ферма имеет решения, если одна из координат P^2, например $z = 0$, а две другие удовлетворяют равенствам $Y^{(q-1)/2} = 1$ и $X^{(q-1)/2} = \beta$. Искомыми решениями являются $y = 1$ и $x = \alpha^{2i+1}$, $i = \overline{0, (q-1)/2 - 1}$. Общее число решений с учетом перестановок координат будет $N = 3(q-1)/2$.

Аналогично получим результат 5. Элементы поля $\alpha^{i(q-1)/6}$ принадлежат мультипликативной группе шестого порядка: $1, \beta^1, \beta^2, \beta^3, \beta^4, \beta^5$, где $\beta = \alpha^{(q-1)/6}$. Мультипликативная группа шестого порядка включает мультипликативные подгруппы вто-

рого порядка $1, \beta^3$ и третьего $1, \beta^2, \beta^4$. По свойству суммы элементов мультипликативной группы имеем $1 + \beta^3 = 0$ и $1 + \beta^2 + \beta^4 = 0$. Первое условие дает $(q-1)/6$ решений для уравнения кривой в виде $z = 0$, $y = 1$ и $x = \alpha^{3(2i+1)}$, $i = \overline{0, (q-1)/6 - 1}$, а с учетом перестановки координат – $3(q-1)/6$ решений. Второе условие – $2(q-1)^2/6$ решений по аналогии с доказательством результата 3. Так как элементы подгрупп не пересекаются и нет других подгрупп порядка 2 или 3, общее число решений $N = (q-1)/2 + (q-1)^2/18$.

Для вывода результата 6 заметим, что элементы поля $\alpha^{i(q-1)/2^m}$ принадлежат мультипликативной группе порядка 2^m: $1, \beta, \beta^2, \ldots, \beta^{2^m-1}$, где $\beta = \alpha^{(q-1)/2^m}$.

Мультипликативная группа содержит мультипликативные подгруппы только четных порядков 2^e, $e = \overline{1, m}$. Если $\left|2^{2^{m-1}}\right| \neq 1(\mathrm{mod} q)$, уравнение Ферма имеет только решения, когда одна из координат, например $z = 0$, а две другие удовлетворяют равенствам $Y^{(q-1)/2^m} = 1$ и $X^{(q-1)/2^m} = \gamma$, где γ есть элемент подгруппы второго порядка. Решением по координате y является значение $y = 1$, а по координате x являются точки $x = \alpha^{2^m i + 2^{m-1}}$, $i = \overline{0, (q-1)/2^m - 1}$. Общее число решений с учетом перестановок координат будет $N = 3(q-1)/2^m$.

Если $2^{2^{m-1}} = -1(\mathrm{mod} q)$, в подгруппе порядка 2^m: $1, \beta, \beta^2, \ldots, \beta^{2^m-1}$, существует элемент $\alpha = 2$, в силу $(2^{2^{m-1}})^2 = 1(\mathrm{mod} q)$. Все степени α порождают подгруппу 2^m порядка, которая является перестановкой элементов подгруппы $1, \beta, \beta^2, \ldots, \beta^{2^m-1}$. Очевидно, что среди элементов подгруппы $1, \alpha, \alpha^2, \ldots, \alpha^{2^m-1}$ будет элемент $\lambda = -2$, так как $\alpha \bullet \alpha^{2^{m-1}} = 2 \bullet 2^{2^{m-1}} = -2(\mathrm{mod} q)$ и элемент $\gamma = (q-1)/2$, в силу $\alpha^{2^{m-1}}/\alpha = 2^{2^{m-1}}/2 = (q-1)/2(\mathrm{mod} q)$. В этом случае кривая Ферма имеет решения, которые определяются условиями: 1) $z = 0$, $Y^{(q-1)/2^m} = 1$, $X^{(q-1)/2^m} = \gamma$; 2) $Z^{(q-1)/2^m} = 1$, $Y^{(q-1)/2^m} = 1$, $X^{(q-1)/2^m} = \lambda$; 3) $Z^{(q-1)/2^m} = 1$, $Y^{(q-1)/2^m} = \gamma$, $X^{(q-1)/2^m} = \gamma$. Число решений по условию 1), как и в предыдущем случае, равно $N_1 = 3(q-1)/2^m$. По условию 2) число корней по каждой координате y и x равно $(q-1)/2^m$, и с учетом перестановок координат общее число решений равно $N_2 = 2(q-1)^2/2^{2m}$. Аналогично для условия 3) с учетом, что перестановок по координатам нет, $N_3 = (q-1)^2/2^{2m}$. Общее число решений определяется суммой $N_1 + N_2 + N_3$, что дает искомое N.

Пример 1. Уравнения $X^m + Y^m + Z^m = 0$ над F_q при $q = 257$ соответствуют случаю, когда делителями порядка поля $q - 1 = 256$ являются степени двойки. Прямое вычисление решений при $m = 128, 64, 32, 16$ дает значения $N = 384, 192, 96, 816$. При $m = 128, 64, 32$ выполняется первое условие для результата 6, т. е. $\left|2^{(q-1)/(2m)}\right| \neq 1(\mathrm{mod} q)$, и число точек кривой равно $N = 3(q-1)/((q-1)/m) = 3m$, что совпадает с прямыми вычислениями. Для $m = 16$ справедливо $2^{(q-1)/(2m)} = -1(\mathrm{mod} q)$ и $N = 3m + 3m^2$, что также совпадает с точным значением.

Точные вычисления числа решений для уравнений Ферма, когда степень уравнения является произвольным делителем порядка конечного поля F_q, является трудоемкой задачей. Ниже рассматриваются оценочные значения для числа точек.

2. ОЦЕНКИ ДЛЯ ЧИСЛА ТОЧЕК КРИВЫХ ФЕРМА В КОНЕЧНОМ ПОЛЕ

Оценки числа решений уравнения Ферма в конечном поле F_q получим на основе вероятностного подхода.

Теорема 1. Пусть кривая $X^m + Y^m + Z^m = 0$ определена над простым полем F_q, где m есть делитель $q - 1$. Оценка для числа точек кривой Ферма при $m > 2$ равна

$$N \approx 2\left\lceil \frac{(q-1)}{2m^2} \right\rceil m^2, \qquad (4)$$

где $\lceil x \rceil$ – округление числа до большего целого.

Доказательство. Решениями уравнения кривой в проективном пространстве P^2 над F_q являются точки $(\gamma : \xi : 1)$, для которых справедливо $\gamma^m + \zeta^m + 1 = 0$ или $\delta + \eta + 1 = 0$, с учетом подстановки $\gamma^m = \delta$, $\zeta^m = \eta$. Пусть $\delta = \beta^i$, $\eta = \beta^j$, где β – образующий элемент поля F_q. Первое условие $\gamma^m + \zeta^m + 1 = 0$ определяет, что степени i и j должны иметь делитель m и β^i, β^j должны удовлетворять второму условию $\delta + \eta + 1 = 0$. Число пар δ, η, удовлетворяющих условию $\delta + \eta + 1 = 0$, равно $(q-1)/2$, а число элементов поля β^i, которые имеют делитель степени m, равно $(q-1)/m$. Образующий элемент $\beta = a$ выбирается из множества чисел $\overline{0, q-1}$, и элементы числового поля вычисляются по правилу $a^i \mathrm{mod} q$. Последнее выражение определяет рандомизатор, для которого соответствие между числовым значением элемента мультипликативной группы и его индексом является псевдослучайным. Таким образом, среднее число пар δ, η, удовлетворяющих условиям 1 и 2, будет равно $n = \frac{(q-1)}{2} \cdot P(i = 0 \mathrm{mod} m, j = 0 \mathrm{mod} m)$, где сомножитель $(q-1)/2$ определяет общее число δ, η, удовлетворяющих условию $\delta + \eta + 1 = 0$, а $P(i = 0 \mathrm{mod} m, j = 0 \mathrm{mod} m)$ – вероятность того, что индексы элементов δ, η имеют делитель m. Предполагая равновероятность распределения индексов элементов поля в парах δ, η, полу-

чим $P(i = 0 \bmod m,\ j = 0 \bmod m) = 1/m^2$ и оценку для n в виде $n = (q-1)/2m^2$. Так как число пар есть целое число, выполним округление n до большего целого. Округление к ближайшему целому для случаев, когда $m > \sqrt{(q-1)}$, привело бы к нулевой оценке числа решений, что не является верным (см. утв. 1). Наконец, учтем перестановку пар по координатам x, y и то, что по каждой координате число корней равно m, получим искомое выражение N. ◊

Рассмотрим следствия данной теоремы, но сначала отметим следующую полезную лемму.

Лемма 1. Пусть кривая $X^m + Y^m + Z^m = 0$ определена над простым полем F_q, где m есть делитель $q-1$. Справедливо следующее:

1) если $(q-1)/m$ содержит делитель 2, в число решений входит $3m$;

2) если $(q-1)/m$ содержит делитель 3, в число решений входит $2m^2$;

3) если $(q-1)/m$ содержит делители 2 и 3, в число решений входит $3m+2m^2$;

4) если $(-2)^{(q-1)/m} = 1 (\bmod q)$, в число решений входит $3m^2$.

Доказательство прямо следует из результатов утверждения 1.

Следствия теоремы 1 с уточнением по результатам леммы 1 представлены в предложении 1.

Предложение 1. Пусть кривая $X^m + Y^m + Z^m = 0$ определена над простым полем F_q, где $m > 2$ есть делитель $q-1$. Оценки числа точек N для кривой Ферма имеют следующий вид.

A) Если $(-2)^{(q-1)/m} \neq 1 (\bmod q)$,

1) $(q-1)/m = 0 (\bmod 2)$, $(q-1)/m \neq 0 (\bmod 3)$, тогда $N \approx 3m + 2\langle (q-1)/(2m^2) \rangle_3 m^2$;

2) $(q-1)/m \neq 0 \bmod 2$, $(q-1)/m = 0 \bmod 3$, тогда $N \approx 2(1 + \langle (q-1)/(2m^2) - 1 \rangle_3) m^2$;

3) $(q-1)/m = 0 \bmod 6$, тогда $N \approx 3m + 2(1 + \langle (q-1)/(2m^2) - 1 \rangle_3) m^2$;

4) $(q-1)/m \neq 0 \bmod 2$, $(q-1)/m \neq 0 \bmod 3$, тогда $N \approx 2\langle (q-1)/(2m^2) \rangle_3 m^2$.

B) Если $(-2)^{(q-1)/m} = 1 (\bmod q)$,

1) $(q-1)/m = 0 (\bmod 2)$, $(q-1)/m \neq 0 (\bmod 3)$, тогда $N \approx 3m + 2\langle (q-1)/(2m^2) - 1 \rangle_3 m^2 + 3m^2$;

2) $(q-1)/m \neq 0 \bmod 2$, $(q-1)/m = 0 \bmod 3$, тогда $N \approx 2(1 + \langle (q-1)/(2m^2) - 2 \rangle_3) m^2 + 3m^2$;

3) $(q-1)/m = 0 \bmod 6$, тогда $N \approx 3m + 2(1 + \langle (q-1)/(2m^2) - 2 \rangle_3) m^2 + 3m^2$;

4) $(q-1)/m \neq 0 \bmod 2$, $(q-1)/m \neq 0 \bmod 3$, $N \approx 2\langle (q-1)/(2m^2) - 1 \rangle_3 m^2 + 3m^2$.

Здесь $\langle x \rangle_3$ – округление x до ближайшего целого, кратного 3.

Доказательство следует из результатов теоремы 1 и леммы 1. Вычисления оценочных значений числа точек и точные вычисления дают хорошее совпадение. Так, для $q = 1021$ имеем совпадение результатов при $m = 510,340,255,204,170,102,85,68,60,51,34,30,20,17,12,5,3,2$ и число точек $N = 1530,231200,765,0,58310,306,14705,9248,0,153,2414,90,800,629,864,965,1008,1022$. Для значений $m = 15,10,6,4$ точные $N = 45,1430,882,1088$, а приближенные $N' = 1335,830,1098,992$. Расхождения проявляются и могут быть существенными, когда степень уравнения становится меньше $\sqrt{q}$. Относительная погрешность вероятностной оценки уменьшается с ростом q.

3. АСИМПОТИЧЕСКИЕ РЕЗУЛЬТАТЫ ПО КРИВЫМ ФЕРМА НАД ПРОСТЫМ ПОЛЕМ

Асимптотические результаты по кривым Ферма над простым полем определяются теоремами 2, 3.

Теорема 2. Асимптотическая граница для отношения максимального числа точек $N_g(q)$ к ее роду g для кривой Ферма в простом поле определяется выражением

$$\limsup_{g \to \infty} \frac{N_g(q)}{g} = 10. \qquad (5)$$

Доказательство. Предел отношения $N_g(q)/g$ определяется максимальной оценкой для числа точек кривых Ферма. Род кривой Ферма равен $g = (m-1)(m-2)/2$, где m – степень уравнения. Движение $g \to \infty$ возможно, когда $q \to \infty$ и $m \to q$. Подставляя в предел отношения $N_g(q)/g$ максимальное значение для числа точек на кривой в простом поле $N \approx 3m + 2(1 + \langle (q-1)/(2m^2) - 2 \rangle_3) m^2 + 3m^2$, что соответствует условию $(-2)^{(q-1)/m} = 1 (\bmod q)$ и $(q-1)/m = 0 (\bmod 6)$ предложения 1, получим искомую оценку (5). ◊

Теорема 3. Асимптотическая граница отношения максимального числа точек $N_g(q)$ для кривой Ферма в простом поле к максимальному числу точек по границе Хассе – Вейля определяется выражением

$$\limsup_{g \to \infty} \frac{N_g(q)}{N_g(q)_{HV}} = \frac{5}{\sqrt{q}}. \qquad (6)$$

Доказательство. Из теоремы 2 следует, что максимальная оценка для числа точек кривых Ферма в простом поле имеет вид $N \approx 3m + 2(1 + \langle (q-1)/(2m^2) - 2 \rangle_3) m^2 + 3m^2$. Выразим m из выражения для рода кривой. Для больших g справед-