НАЦІОНАЛЬНА АКАДЕМІЯ НАУК УКРАЇНИ
ІНСТИТУТ ПРОБЛЕМ МОДЕЛЮВАННЯ В ЕНЕРГЕТИЦІ ім. Г.Є. ПУХОВА

ГІЛЬГУРТ СЕРГІЙ ЯКОВИЧ

УДК 004.274:004.056

МЕТОДИ ТА ЗАСОБИ СТВОРЕННЯ РЕКОНФІГУРОВНИХ СИГНАТУРНИХ ЗАСОБІВ ЗАХИСТУ ІНФОРМАЦІЇ КОМП'ЮТЕРНИХ СИСТЕМ І МЕРЕЖ

05.13.05 – комп'ютерні системи та компоненти

Автореферат

дисертації на здобуття наукового ступеня
доктора технічних наук

Київ – 2020

Дисертацією є рукопис.

Робота виконана в Інституті проблем моделювання в енергетиці ім. Г.Є. Пухова НАН України, м. Київ.

Науковий консультант доктор технічних наук,
старший науковий співробітник
Чемерис Олександр Анатолійович,
Інститут проблем моделювання в енергетиці
ім. Г.Є. Пухова НАН України,
заступник директора з наукової роботи.

Офіційні опоненти: доктор технічних наук, професор
Лужецький Володимир Андрійович,
Вінницький національний технічний університет,
завідувач кафедри захисту інформації;

доктор технічних наук, професор
Опанасенко Володимир Миколайович,
Інститут кібернетики імені В.М. Глушкова
НАН України, провідний науковий співробітник
відділу мікропроцесорної техніки;

доктор технічних наук, доцент
Сергієнко Анатолій Михайлович,
Національний технічний університет України
«Київський політехнічний інститут
імені Ігоря Сікорського», професор
кафедри обчислювальної техніки.

Захист відбудеться «11» грудня 2020 р. о 14 годині на засіданні спеціалізованої вченої ради Д 26.185.01 Інституту проблем моделювання в енергетиці ім. Г.Є. Пухова НАН України за адресою: 03164, м. Київ, вул. Генерала Наумова, 15.

З дисертацією можна ознайомитися в бібліотеці Інституту проблем моделювання в енергетиці ім. Г.Є. Пухова НАН України, м. Київ за адресою: 03164, м. Київ, вул. Генерала Наумова, 15.

Автореферат розісланий « 9 » листопада 2020 р.

Вчений секретар
спеціалізованої вченої ради В.В. Душеба

ЗАГАЛЬНА ХАРАКТЕРИСТИКА РОБОТИ

Актуальність теми. Загрози інформаційної безпеки відносяться до числа найважливіших категорій проблем, що стоять сьогодні перед людством. Для протидії ним використовуються різні заходи – юридичні, адміністративні, організаційні, технічні, криптографічні тощо. Кожному напряму захисту притаманні свої складності та проблеми. Серед технічних засобів захисту інформації важливу роль відіграють системи, принцип дії яких заснований на пошуку в інформаційних потоках заздалегідь відомих ознак шкідливої активності – так званих сигнатур. На відміну від альтернативних підходів, які також активно та успішно розвиваються останнім часом, сигнатурні засоби досі демонструють кращі результати щодо точності виявлення зловживань, генерують менше помилок розпізнавання як першого, так и другого роду.

Сучасні сигнатурні засоби технічного захисту, таки як мережеві системи виявлення вторгнень (МСВВ), противірусні сканери, фільтри спаму, засоби протидії мережевим хробакам тощо, мають вирішувати в реальному часі обчислювально складну задачу множинного розпізнавання рядків. У зв'язку з припиненням зростання частоти традиційних процесорів, а також через сталий зріст зловмисної активності програмні додатки вже не забезпечують потрібної продуктивності. Тому останнім часом розробники вимушені звертатися до апаратних рішень, найчастіше – на базі програмованих логічних інтегральних схем (ПЛІС). Висока продуктивність програмованої логіки в поєднанні з гнучкістю, близької до програмної, природним чином відповідає складній та динамічній сутності завдань інформаційного захисту. Локальний характер задач захисту інформації обумовлює використання в якості платформи для таких рішень реконфігуровних обчислювачів (прискорювачів), які містять мікросхему ПЛІС, пристрої оперативної пам'яті, допоміжні компоненти, а також мають уніфіковані засоби взаємодії з традиційною комп’ютерною технікою.

Але використання реконфігуровних засобів само по собі не вирішує проблеми браку продуктивності. Збільшення об’ємів мережевого трафіку триває. Складність та витонченість зловмисної активності зростають. Також з часом змінюється коло об'єктів, що наражаються на небезпеку: якщо сьогодні зовнішні атаки загрожують переважно локальним мережам компаній та організацій, то згодом, по мірі застосування інформаційних технологій інтернету речей та кіберфізичних систем, до зони ризику почнуть підпадати новітні виробничі підприємства.

Розробкою та вдосконаленням реконфігуровних апаратних сигнатурних засобів технічного захисту інформації (**РАСЗТЗІ**) комп’ютерних систем і мереж сьогодні в світі займаються численні дослідники. Запропоновано багато технічних рішень та їх модифікацій, заснованих на різних за своєю природою принципах та підходах до побудови швидкісних цифрових схем розпізнавання патернів (фіксованих послідовностей символів) у вхідному потоці даних. Кожному з підходів притаманні як переваги, так і недоліки. Проте, досі не виявлено пануючого підходу, який за більшістю показників випереджав би інші. Комбінування тобто об’єднання в єдиному пристрої різних підходів таким чином, щоб їх переваги посилювалися, а недоліки нівелювалися, вже зараз застосовується дослідниками при розробці апаратних засобів захисту, але, як правило, евристично, без наукового підґрунтя.

Тому виникає актуальна науково-технічна проблема, що потребує вирішення, яка може бути сформульована як задача розробки та розвитку методів побудови комбінованих обчислювальних структур для підвищення ефективності реконфігуровних сигнатурних технічних засобів захисту інформації.

Дослідженнями щодо використання ПЛІС в сигнатурних системах захисту інформації в світі займається багато вчених. Починали цю роботу науковці з США – J.W. Lockwood, S. Dharmapurikar, V.K. Prasanna Z.K. Baker, C.R. Clark, D.E. Schimmel, Y.H. Cho, W.H. Mangione-Smith, H. Chen, Y. Chen, D.H. Summerville та ін.; продовжували вчені з Греції, Нідерландів, Великої Британії, Швейцарії та інших країн Європи – S. Vassiliadis, I. Sourdis, D.N. Pnevmatikatos, K.G. Anagnostakis, E.P. Markatos, T. Koçak, J. Lunteren та ін.; згодом долучилися азійські вчені з Індії, Ірану, В'єтнаму, Кореї та Японії, а також численні дослідники з Китаю.

В СРСР та на пострадянському просторі проблемами реконфігуровних обчислень активно займалися такі вчені, як В.І. Варшавський, Е.В. Євреінов, А.В. Каляєв, І.І. Лєвін та ін., в Україні – В.М. Глушков, О.В. Палагин, В.М. Опанасенко, А.О. Мельник, А.М. Сергієнко, В.І. Хаханов, В.Ф. Євдокимов, В.В. Мохор, О.А. Чемерис та багато інших. Але саме застосуванню програмованої логіки для вирішення ресурсоємних задач сигнатурного аналізу в галузі захисту інформації вітчизняними спеціалістами суттєвої уваги досі не приділялося.

Зв’язок роботи з науковими програмами, планами, темами. Дослідження, що проводились при виконанні дисертаційної роботи, здійснювалися у відповідності з планом науково-дослідних робіт Інституту проблем моделювання в енергетиці ім. Г.Є. Пухова НАН України в рамках наступних науково-дослідних робіт:

«Дослідження та розробка методів оцінювання захищеності інформації в розподілених високопродуктивних інформаційних системах при вирішенні задач енергетики (шифр МОД-Д)», що виконувалась у 2014-2018 рр. (номер державної реєстрації 0114U002361) – створено методологію побудови систем технічного захисту інформації на базі реконфігуровних обчислювачів;

«Розвиток теорії, розробка новітніх інформаційних технологій в задачах комплексного моделювання та управління процесами перетворення та використання енергії (шифр НОВІНТЕХ)», що виконується з 2017 року по теперішній час (номер державної реєстрації 0117U004347) – запропоновані підходи щодо використання хмарних та грід-ресурсів для синтезу конфігурацій апаратних засобів інформаційної безпеки, здійснено дослідження та вибір апаратної платформи для реалізації сигнатурних методів технічного захисту інформації на об'єктах енергетиці;

«Підтримка та розвиток грід-сайту Інституту проблем моделювання в енергетиці ім. Г.Є. Пухова НАНУ, як ресурсного центра NGI_UA, та створення грід-сервіса централізованого синтезу конфігурацій для апаратних прискорювачів задач інформаційної безпеки в енергетичній галузі (шифр ГРІДІПМЕМОН-15)», що виконувалась у 2015 р. згідно Цільової комплексної програми наукових досліджень НАН України «Грід-інфраструктура і грід-технології для наукових і науково-прикладних застосувань» (номер державної реєстрації 0115U002876) – запропоновано принципи централізованого синтезу конфігурацій для реконфігуровних прискорювачів задач інформаційної безпеки;

«Підтримка та розвиток грід-сайту Інституту проблем моделювання в

енергетиці ім. Г.Є. Пухова НАНУ та створення системи централізованого програмування реконфігуровних прискорювачів задач інформаційної безпеки в енергетичній галузі (шифр ГРІДІПМЕМОН-16)», що виконувалась у 2016 р. згідно Цільової комплексної програми наукових досліджень НАН України «Грід-інфраструктура і грід-технології для наукових і науково-прикладних застосувань» (номер державної реєстрації 0116U006907) – створено принципи використання грід-системи для централізованого синтезу реконфігуровних систем захисту інформації;

«Підтримка грід-сайту Інституту проблем моделювання в енергетиці ім. Г.Є. Пухова НАНУ та використання хмарної інфраструктури для централізованого програмування реконфігуровних засобів інформаційної безпеки в енергетичній галузі (шифр ГРІДІПМЕМОН-18)», що виконувалась у 2018 р. згідно Цільової комплексної програми наукових досліджень НАН України «Грід-інфраструктура і грід-технології для наукових і науково-прикладних застосувань» (номер державної реєстрації 0118U001370) – створено принципи використання хмарних обчислень для синтезу реконфігуровних систем захисту інформації.

«Підтримка грід-сайту ІПМЕ ім. Г.Є. Пухова НАН України та модернізація веб-сервісу централізованого програмування реконфігуровних засобів інформаційної безпеки на базі гріду та хмарної інфраструктури (шифр ГРІДІПМЕМОН-19)», що виконувалась у 2019 року згідно Програми інформатизації НАН України на 2019 р. (номер державної реєстрації 0119U001812) – розроблено методи створення ефективних структур сигнатурного розпізнавання на базі реконфігуровних прискорювачів для інформаційної безпеки.

Мета і завдання дослідження. Метою дисертаційної роботи є підвищення ефективності реконфігуровних сигнатурних технічних засобів захисту інформації в комп'ютерних системах і мережах шляхом постановки та вирішення задачі комбінування різних підходів до побудови схем розпізнавання таких засобів.

Для досягнення поставленої мети в дослідженні було потрібно вирішити наступні задачі:

1. Проаналізувати проблеми технічного захисту інформації, які вирішуються сигнатурними засобами.
2. Дослідити технічні можливості сучасних ПЛІС та пристроїв на їх основі як апаратної бази для вирішення задач технічного захисту інформації.
3. Проаналізувати принципи функціонування та побудови РАСЗТЗІ як технічних систем. Сформулювати та класифікувати показники їх ефективності.
4. Виокремити ефективні підходи та їх модифікації до побудови модулів розпізнавання РАСЗТЗІ, на структурному рівні виявити та формалізувати властивості кожного з них, враховуючи специфіку реалізації реконфігуровними засобами.
5. Розробити методи прискореного виконання процедур оптимізації для використання в методах комбінування з метою зменшення їх обчислювальної складності до прийнятних значень.
6. Застосувати розроблені методи прискореного обчислення для виявлення функціональних залежностей кількісних показників ефективності компонентів РАСЗТЗІ, побудованих за обраними підходами та їх модифікаціями, від характеристик використаного РУО та властивостей набору патернів, що мають

розпізнаватися.

7. Сформулювати та вдосконалити методи та принципи комбінування в одному пристрої декількох різних підходів із застосуванням оптимізаційних методів для використання переваг кожного з них.
8. Розробити алгоритми реалізації вдосконалених методів комбінування.
9. Розробити програмні засоби для перевірки та аналізу запропонованих методів і засобів, провести обчислювальні експерименти щодо їх кількісної оцінки.
10. Дослідити та розвити принцип централізованого створення реконфігуровних обчислювачів з використанням високопродуктивних розподілених та хмарних середовищ і застосувати його для реалізації запропонованих методів та засобів.

Об'єктом дослідження є процеси технічного захисту інформації в комп'ютерних системах і мережах.

Предметом дослідження є реконфігуровні апаратні сигнатурні засоби технічного захисту інформації в комп'ютерних системах і мережах.

Методи дослідження. При виконанні поставлених у дисертаційній роботі задач були використані загальнонаукові та спеціальні методи досліджень: положення теорії обчислювальних систем, апарат булевої алгебри, теорія автоматів, теорія обчислень на рядках, елементи теорій графів, множин, алгоритмів та комбінаторики.

Перевірка наукових положень та методів проводилася шляхом обчислювальних розрахунків за розробленими алгоритмами с використанням експериментальних програмних засобів а також шляхом прогонів тестових завдань на макетному зразку веб-сервісу централізованого створення реконфігуровних засобів інформаційної безпеки STRAGS, побудованого на базі кластера Інституту проблем моделювання в енергетиці ім. Г.Є. Пухова НАН України.

Наукова новизна одержаних результатів. У дисертаційній роботі вирішена науково-технічна проблема, що полягає в розробці та розвитку методів побудови комбінованих обчислювальних структур для підвищення ефективності реконфігуровних сигнатурних технічних засобів захисту інформації. Наукова новизна одержаних результатів полягає у наступному:

1. Вперше розроблено метод прискореного обчислення технічних характеристик компонентів РАСЗТЗІ, що використовуються в якості цільових функцій процедур оптимізації, який на відміну від відомих методів проектування реконфігуровних пристроїв не потребує виконання витратних за часом процедур синтезу цифрових схем, дозволяючи за рахунок формування математичного опису здійснювати швидку оцінку та порівняння обчислювальних структур за визначеними показниками ефективності.

2. Вперше розроблено метод прискорення процедури оптимізації паралельного комбінування, який відрізняється впорядкуванням патернів в наборі за певним параметром, що дозволяє замість повного перебору комбінаторно великої кількості варіантів здійснювати поділ патернів між блоками розпізнавання, що комбінуються, за лінійним законом часу, послідовно змінюючи параметр впорядкування. Метод не гарантує знаходження глобального оптимуму, проте зменшує обчислювальну складність процесу оптимізації до прийнятних для практичного використання значень.

3. Вперше сформульовано та вдосконалено метод паралельного комбінування,

який за рахунок паралельного з'єднання різних за принципами побудови блоків розпізнавання дозволяє підвищити показники ефективності модулю розпізнавання РАСЗТЗІ. Запропоноване вдосконалення відрізняється застосуванням розроблених методів прискорення, що уможливлює використання процедур оптимізації за рахунок зниження часових витрат до практично прийнятних значень.

4. Вперше сформульовано та вдосконалено метод послідовного каскадування, який за рахунок послідовного з'єднання різних за принципами побудови блоків розпізнавання та уточнення попереднього розпізнавання, використовуючи процедуру оптимізації поділу між ними питомих патернів по довжині, дозволяє підвищити показники ефективності модулю розпізнавання РАСЗТЗІ. Запропоноване вдосконалення відрізняється застосуванням розробленого методу прискореного обчислення технічних характеристик, що уможливлює використання процедур оптимізації за рахунок зниження часових витрат до прийнятних для практичного використання значень.

5. Отримав подальший розвиток метод вертикального об’єднання, який за рахунок тісного сполучання в одному блоці кількох підходів або технічних рішень дозволяє підвищити показники ефективності модулю розпізнавання РАСЗТЗІ. Запропонований розвиток відрізняється використанням багатовимірної таблиці сумісності, що дозволяє формалізувати наявний досвід численних дослідників та спростити процедуру оптимізації вибору найбільш ефективної за наданих умов комбінації задіяних даним методом підходів або технічних рішень.

6. Вперше сформульовано та вдосконалено принцип комбінування методів комбінування, який за рахунок ієрархічного використання сформульованих, вдосконалених та розвитих методів комбінування дозволяє підвищити показники ефективності модулю розпізнавання РАСЗТЗІ до значень, недосяжних при використанні кожного з методів комбінування окремо. Запропоноване вдосконалення відрізняється застосуванням розроблених методів прискорення, що уможливлює використання загальної процедури оптимізації в зв’язку зі зниженням часу розрахунків до прийнятних для практичного використання значень.

Практичне значення одержаних результатів полягає в тому, що застосування запропонованих методів побудови РАСЗТЗІ дозволить розробникам створювати більш ефективні засоби захисту, які використовуватимуться для протидії зовнішнім та внутрішнім атакам на комп’ютерні системи та мережі. Здобуті результати щодо підвищення ефективності вирішення задачі множинного розпізнавання рядків актуальні й для мережевих застосувань, не пов’язаних з захистом інформації: прискореної обробки XML-запитів, класифікації пакетів з відновленням з'єднань для управління технологією QoS, фільтрації IP-телефонії, вимірювання трафіку, оптимізації кешування та ін. Запропоновані методи також можуть бути застосовані для вирішення важливих проблем з інших галузей науки, зокрема, інтелектуального пошуку даних (data mining), аналізу молекул ДНК тощо.

Результати дисертаційної роботи впроваджено в Національному аерокосмічному університеті ім. Н.Є Жуковського "Харківський авіаційний інститут", на приватному підприємстві "Геракс", в Національному авіаційному університеті та в Київському державному університеті ім. Т.Г. Шевченка, що підтверджується відповідними документами.

Особистий внесок здобувача. Всі основні положення й результати дисертаційної роботи, що виносяться на захист, отримано автором самостійно. У роботах, які опубліковані в співавторстві, здобувачеві належать наступні результати: [4, 18] – аналіз можливостей сучасних реконфігуровних обчислювачів в якості апаратної бази для вирішення задач інформаційної безпеки; [7, 8, 12, 39] – аналіз задач технічного захисту, які вирішуються сигнатурними реконфігуровними системами; [10, 48] – узагальнена структура РАСЗТЗІ; [11] – аналіз задачі розпізнавання рядків як обчислювально складного процесу та шляхів її вирішення реконфігуровними засобами; [16, 17] – аналіз можливостей використання некомерційних МСВВ та їх баз даних сигнатур для оцінки ефективності РАСЗТЗІ; [20] – аналіз властивостей МСВВ щодо протидії цілеспрямованим атакам на них в якості функціонального показника ефективності РАСЗТЗІ; [23, 25, 26] – концепція та методологія централізованого синтезу реконфігуровних обчислювачів для вирішення задач захисту інформації; [33] – структура та склад реконфігуровного обчислювача; [51] – методи підвищення ефективності РАСЗТЗІ шляхом комбінування; [52] – аналіз та вибір елементної бази для реалізації апаратно-програмного комплексу.

Апробація результатів дисертації. Основні теоретичні та прикладні положення і висновки дисертаційного дослідження представлялися та обговорювалися на 16 міжнародних та всеукраїнських науково-технічних та науково-практичних конференціях і семінарах: Міжнародна конференція «Параллельные вычисления и задачи управления», *м. Москва, РФ*, 2004, 2006 рр.; Міжнародна науково-технічна конференція «ІТУЕС», *м. Київ, Україна*, 2005 р.; Міжнародна науково-технічна конференція «Искусственный интеллект. Интеллектуальные и многопроцессорные системы», *сел. Кацівелі, Україна*, 2006, 2011 рр.; Міжнародна науково-практична конференція «Інтелектуальні системи прийняття рішень та інформаційні технології», *м. Чернівці, Україна*, 2006 р.; Міжнародна наукова конференція «Параллельные вычислительные технологии», *м. Санкт-Петербург, РФ*, 2008 р.; Міжнародна науково-технічна конференція «Многопроцессорные вычислительные и управляющие системы», *с. Дивноморское, РФ,* 2009 р.; Міжнародна наукова конференція «Моделювання», *м. Київ, Україна*, 2010, 2016, 2018 рр.; Міжнародна науково-технічна конференція «Сучасні комп'ютерні системи та мережі: розробка та використання», *м. Львів, Україна*, 2011 р.; Міжнародна науково-практична конференція «Сучасні проблеми і досягнення в галузі радіотехніки, телекомунікацій та інформаційних технологій», *м. Запоріжжя, Україна*, 2014, 2018 рр.; Міжнародна науково-технічна конференція «Високопродуктивні обчислення» (HPC-UA 2018), *м. Київ, Україна*, 2018 р.; IV всеукраїнська науково-технічна конференція «Перспективні напрями захисту інформації», *м. Одеса, Україна*, 2018 р.; IX міжнародна наукова конференція «Безпека інформаційних технологій» (ITSec-2019), *м. Шарм-ель-Шейх, Єгипет*, 2019 р.; Республіканська науково-практична конференція «Цифровые технологии в промышленности», *м. Актау, Казахстан*, 2019 р.; VII міжнародна наукова конференція «Захист інформації і безпека інформаційних систем», *м. Львів, Україна*, 2019 р.; Науково-практична конференція «Кібербезпека енергетики», *м. Одеса, Україна*, 2019 р.; V Всеукраїнська науково-практична конференція «Перспективні

напрями захисту інформації 2019», *смт Затока Одеської обл., Україна*, 2019 р.

Публікації. Наукові положення, висновки і рекомендації дисертаційного дослідження опубліковані в 52 наукових роботах, які відповідають вимогам до опублікування результатів дисертацій, у тому числі: 30 – у наукових фахових журналах та збірниках наукових праць, з яких 6 – у наукових журналах, що індексуються міжнародними наукометричними базами даних, 21 публікація у працях і матеріалах наукових конференцій, один патент, 31 публікацію підготовлено одноосібно, англійською мовою – 3 публікації.

Структура та обсяг дисертації. Дисертаційна робота складається з анотації, вступу, семи розділів, висновків, списку використаних джерел та додатків. Загальний обсяг роботи становить 383 сторінки, із них основного тексту дисертації – 298 сторінок, 62 рисунки, 10 таблиць, список використаних джерел включає 227 найменувань та займає 25 сторінок, обсяг додатків складає 27 сторінок.

ОСНОВНИЙ ЗМІСТ РОБОТИ

У вступі наведено загальну характеристику дисертаційної роботи, обґрунтовано актуальність та важливість теми дослідження, сформульовано його мету та задачі, окреслено об’єкт і предмет, наведено відомості про зв'язок роботи з науковими програмами, планами, темами, визначено наукову новизну та практичну цінність роботи. Також у вступі наведено дані про особистий внесок здобувача в оприлюднених у співавторстві наукових працях, відомості про апробацію результатів роботи, надано інформацію про структуру дисертації.

У першому розділі проаналізовано, з одного боку, проблеми технічного захисту інформації, які вирішуються сигнатурними засобами, з іншого – технічні можливості сучасних ПЛІС як апаратної бази для вирішення складних обчислювальних задач інформаційного захисту.

Розглянуто фундаментальні поняття інформаційної безпеки та захисту інформації, місце, яке займає в ієрархії заходів забезпечення інформаційної безпеки рівень технічного захисту інформації. З відомих засобів виокремлено сигнатурні засоби технічного захисту інформації (СЗТЗІ), функціонування яких засновано на використанні сигнатур – описів конкретних загроз інформаційної безпеки.

Головний недолік СЗТЗІ полягає в тому, що вони не здатні розпізнавати нові атаки, для яких ще не створені сигнатури. Тому останнім часом активно і успішно розвиваються такі несигнатурні підходи та класи підходів, як машинне навчання, добування даних – data mining і використання баз знань, а також статистичний, поведінковий та евристичний підходи. Але, не зважаючи на перспективність таких напрямів, рішення подібного плану досі потерпають від відносно високої інтенсивності помилок розпізнавання (як першого, так і другого роду), складності процесу налаштування, довготривалості навчання та створення профілю нормального стану. Тому в даному дослідженні розглянуті саме СЗТЗІ, до яких належать сигнатурні мережеві системи виявлення вторгнень – МСВВ, противірусні сканери, фільтри спаму та засоби протидії мережевим хробакам.

Історично першими СЗТЗІ, які широко використовували реконфігуровні пристрої, були МСВВ. Як наслідок, ці засоби найбільш вивчені та пророблені. Тому

в даній роботі, не втрачаючи загальності, СЗТЗІ досліджуються на прикладі МСВВ.

Для роботи з СЗТЗІ потрібні певні визначення. *Сигнатура* в даному дослідженні розглядається як сукупність інформації про конкретну загрозу, що зберігається в базі даних сигнатур відповідного СЗТЗІ. Під терміном *патерн* розуміється зразок текстового рядка (фіксована послідовність символів в певному кодуванні), який входить до складу сигнатури і відшукується у вхідних даних. *Словник патернів* – вся сукупність патернів, що міситься у базі даних сигнатур. *Набір патернів* – патерни, що міситься тільки в деяких сигнатурах бази даних МСВВ і обрані з певною метою. Термін *множина патернів* використовується для позначення абстрактної низки патернів в математичному сенсі, не акцентуючи увагу на застосуванні. *Ефект самоподоби* – явище, яке полягає в частковому взаємному дублюванні фрагментів патернів, коли початок одного патерну (префікс), середня частина (інфікс) або кінець (суфікс) збігається з деякою частиною іншого. Ключовою особливістю СЗТЗІ є необхідність вирішувати в реальному часі складну обчислювальну *задачу множинного розпізнавання патернів* (ЗМРП). Наведено математичну постановку ЗМРП в термінах теорії обчислень на рядках, яку деталізовано для технічної реалізації.

Стале зростання обчислювальної складності ЗМРП призвело до того, що програмні рішення на традиційних процесорних засобах вже не встигають вчасно її вирішувати. Одним із шляхів вирішення проблеми є перехід до апаратних рішень. Внаслідок локальної природи задач СЗТЗІ виконання ЗМРП на високопродуктивних засобах, таких як кластери, грід або хмарне середовище, не є ефективним. Більш прийнятним виявляється використання приєднаних обчислювачів (прискорювачів), які розташовуються безпосередньо в місці вирішення задачі, тому не потребують швидкого пересилання великих обсягів даних на значну відстань.

Автором проведено порівняльний аналіз застосування в якості апаратної платформи для реалізації апаратних СЗТЗІ таких засобів як спеціалізовані сопроцесори, нейромережеві прискорювачі, мережеві процесори, прискорювачі на базі багатоядерних процесорів (Multi-Core), графічні акселератори, трійкова асоціативна пам'ять (TCAM) та реконфігуровні уніфіковані обчислювачі (РУО) на базі ПЛІС. Результати аналізу призводять до висновку, що складній та динамічній природі задач інформаційного захисту краще відповідає реконфігуровна елементна база. Тому предметом даного дослідження було обрано саме реконфігуровні апаратні сигнатурні засоби технічного захисту інформації – **РАСЗТЗІ**.

Програмовані логічні інтегральні схеми вже тривалий час успішно використовуються в техніці. Методи, що запропоновані в даному дослідженні, використовують деякі їх особливості, тому в розділі наведено потрібні відомості про їх внутрішню організацію. Узагальнюючі, можна вважати, що в найбільш поширених ПЛІС в якості основних обчислювальних ресурсів використовуються: так звані LUT (lookup table) тобто *логічні таблиці* (ЛТ), *тригери*, які разом з ЛТ входять до складу логічних комірок (logic cell), а також ресурси *блокової пам'яті* у вигляді блоків BRAM.

Формою реалізації реконфігуровної апаратної бази у вигляді приєднаних обчислювачів є РУО, які містять щонайменш одну мікросхему ПЛІС та бортовий (on-board) оперативний запам'ятовуючий пристрій (ОЗП). *Пам'ять бортового ОЗП* є

четвертим типом ресурсів поряд з ЛТ, тригерами та BRAM. В розділі також з'ясовано, що, внаслідок більш тривалого, ніж у традиційної комп'ютерної техніки, терміну морального старіння РУО, в експлуатації одночасно знаходяться дуже різні за можливостями прискорювачі, параметри яких відрізняються на декілька порядків.

У другому розділі досліджено принципи створення та функціонування РАСЗТЗІ, сформульовано показники їх ефективності, проведений попередній аналіз відомих підходів до побудови РАСЗТЗІ.

На рис. 1. наведено узагальнену структуру головної складової МСВВ – аналізатора, отриману автором в результаті вивчення та аналізу численних розробок, в яких даний компонент реалізовано на ПЛІС.

Ключовим компонентом будь-якого РАСЗТЗІ є модуль розпізнавання (**МР**). Оскільки саме в ньому вирішується обчислювально складна задача ЗМРП, його характеристики безпосередньо впливають на властивості СЗТЗІ в цілому. Тому в подальшому дослідженні основна увага приділяється саме МР РАСЗТЗІ.

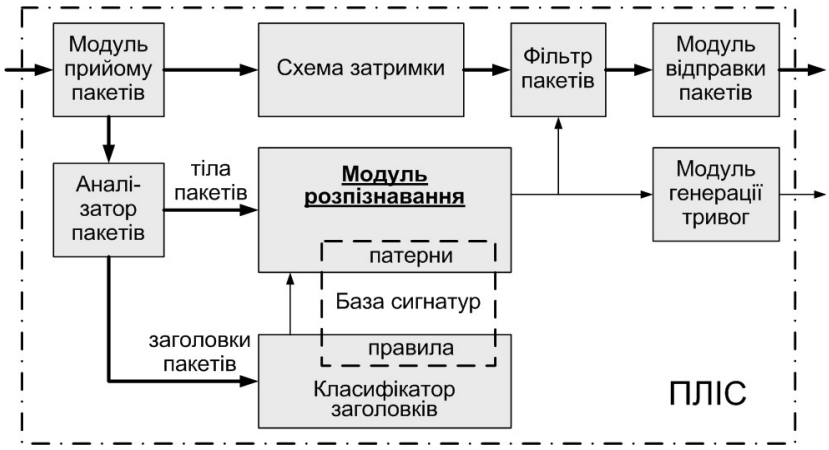


Рис. 1. Узагальнена структурна схема аналізатора МСВВ на базі ПЛІС

Специфічною особливістю застосування реконфігуровних засобів для апаратного прискорення роботи СЗТЗІ є необхідність повторного синтезу деяких компонентів, що регулярно виникає під час їх функціонування. Назвемо таку операцію *процедурою оперативного оновлення* (ПОО).

Причиною потреби в подібній процедурі може бути, по-перше, поява описів нових атак, що не були відомі раніше, по-друге – зміна умов роботи інформаційної системи, що захищається (модифікація локальної мережі, зміна її складу або структури, модифікація програмного забезпечення тощо). У зв'язку з тим, що в процесі оперативного оновлення змінюється лише частка обладнання (яке безпосередньо пов'язане з обробкою сигнатур та патернів, що в неї містяться – МР та класифікатор заголовків), виконання повного циклу створення цифрової схеми в ПЛІС не потрібно. Більшу частину роботи можна виконати завчасно та уникнути під час проведення ПОО, що прискорює процес повторного синтезу аналізатора. (Назвемо відповідні компоненти *змінною* та *постійною складовими* РАСЗТЗІ). Тоді процедура ПОО зводиться до синтезу цифрової схеми розпізнавання в два етапи. На першому етапі згідно питомого набору патернів формуються обчислювальні структури оновлених компонентів у вигляді тексту на мові опису апаратури. На другому – отримані описи в автоматичному режимі перетворюються на *конфігурацію* для ПЛІС – набір бітів у вигляді бінарного файлу (bitstream), який, завантажуючись в програмовану мікросхему, формує її внутрішню структуру.

Щоб мати змогу оцінювати властивості розробок сигнатурних засобів захисту та їх компонентів і порівнювати різні підходи до побудови РАСЗТЗІ та окремі технічні рішення, потрібно визначитися з критеріями їх ефективності та

відповідними показниками. Аналіз світового досвіду розробок РАСЗТЗІ дозолив автору виявити наступну ієрархію їх показників ефективності (ПЕ).

Всі показники можна поділити на три категорії: *основні*, *проміжні* (що пов'язують деякі з основних) та *похідні* (що формуються з кількох інших).

До категорії *основних* ПЕ належать: вартісні показники, показники продуктивності (швидкісні показники) та функціональні показники.

До вартісних ПЕ належать: обсяги *логічних ресурсів* програмованої логіки, які задіяні для створення цифрової схеми, *витрати на пам'ять* (як зовнішню відносно кристалу ПЛІС, так і внутрішню – блочну пам'ять BRAM та розподілену у вигляді тригерів логічних комірок), а також *інші витрати*, що можуть бути оцінені в вартісних одиницях.

До ПЕ продуктивності відносяться, з одного боку – *об'єм словнику патернів* (кількість різних патернів, що розпізнає засіб), з іншого – *швидкодію* засобу, яка характеризується або часом обробки даних пристроєм, або пропускною здатністю. Важливим якісним швидкісним ПЕ РАСЗТЗІ є *передбачуваність пропускної здатності*, яка спрощує оперування іншими швидкісними характеристиками системи та полегшує інтеграцію МР з рештою компонентів технічної системи.

До функціональних ПЕ відносяться переважно якісні показники, а саме: спроможність протидіяти атакам, що спрямовані на РАСЗТЗІ, здатність до оновлення словнику патернів без припинення процесу розпізнавання (властивість динамічної реконфігурації), селективне розпізнавання (зміна підмножини патернів за зовнішнім сигналом), здатність працювати у режимі запобігання вторгнень (для МСВВ) та ін. При створенні деяких РАСЗТЗІ висуваються спеціальні вимоги до їх здібностей, в результаті чого з'являються додаткові функціональні показники.

До проміжних ПЕ належать масштабованість (трьох типів: за пропускною здатністю, за об'ємом словнику патернів і за довжиною патернів) та здатність використання надлишковості словнику патернів (з метою покращення швидкісних і ресурсних характеристик).

Похідні ПЕ мають вигляд певної функціональної залежності (математичного виразу довільної складності) від декількох кількісних ПЕ. Потреба в таких показниках з'явилася в результаті накопичення досвіду створення та використання РАСЗТЗІ світовою дослідницькою спільнотою для більш витонченого вимірювання та порівняння кількісних характеристик технічних рішень для засобів захисту.

В результаті попереднього аналізу існуючих розробок РАСЗТЗІ автором з'ясовано, що при побудові апаратних схем множинного розпізнавання найкращі здібності в сенсі ефективності виявили наступні три підходи (та їх модифікації):

– асоціативна пам'ять на базі цифрових компараторів;
– фільтр Блума на базі геш-функцій;
– алгоритм Ахо–Корасік на базі скінченних автоматів.

У третьому розділі для кожного з цих напрямів автором запропоновано деталізовану формалізацію властивостей та особливостей, без чого неможливо успішно здійснювати їх комбінування для створення ефективних МР РАСЗТЗІ. При цьому враховувалися переваги та недоліки кожного підходу в сенсі сформульованих раніше ПЕ, можливості покращення показників, складнощі реалізації на ПЛІС, які при цьому виникають і шляхи їх подолання.

Підхід на основі асоціативної пам'яті та цифрових компараторів.

Асоціативна пам'ять (АП) – Content Addressable Memory (CAM), є класом пристроїв, що створювалися саме для швидкого розпізнавання кодів і виконують функцію, протилежну традиційному ОЗП: за змістом відшукують місце розташування даних в запам'ятовуючому пристрої або сигналізують про їх відсутність. Швидкодіючою основою АП на ПЛІС є цифрові компаратори (ЦК).

Безпосереднім рішенням, щодо виявлення збігу вхідних символів з патернами є набір цифрових компараторів, кожен з яких порівнює вхідний байт з наперед заданим символом. Назвемо таку схему *Базовою схемою на АП* або схемою *BsCAM* (рис. 2). Після подачі вхідних даних схема BsCAM здатна отримувати результат на виході за один такт синхронізації. Її перевагами є також простота та регулярність структури. Головний недолік – значне споживання логічних ресурсів ПЛІС.

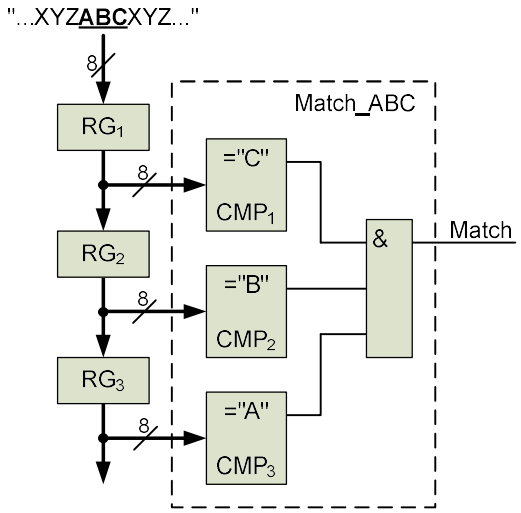


Рис. 2. Безпосереднє розпізнавання цифровими компараторами

Реалізація реконфігуровними засобами ускладнюється, по-перше, перенавантаженням виходів регістрів, побудованих на штатних компонентах ПЛІС, сигнали з яких додаються на велику кількість компараторів, по-друге, необхідністю об’єднувати велику кількість бітових сигналів багатовходовою схемою "І" для довгих патернів. Обидві проблеми вирішуються побудовою конвеєрних схем розгалуження, що призводить до додаткового зростання апаратних витрат, тобто є причиною поганої масштабованості підходу за об’ємом словнику патернів.

Скоротити апаратні витрати АП дозволяє повторне використання компараторів, що в граничному випадку призводить до виділення лише одного ЦК на кожний символ алфавіту, а їх комбінації формуються за допомогою цифрових схем затримки (ЦСЗ) та схем "І", як показано на рис. 3.

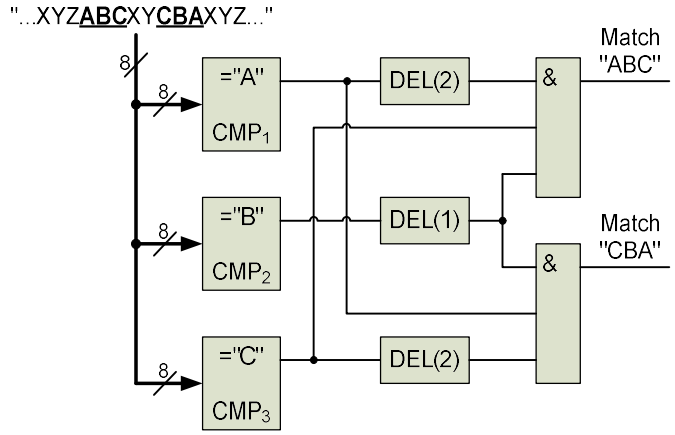


Рис. 3. Схемне рішення DCAM

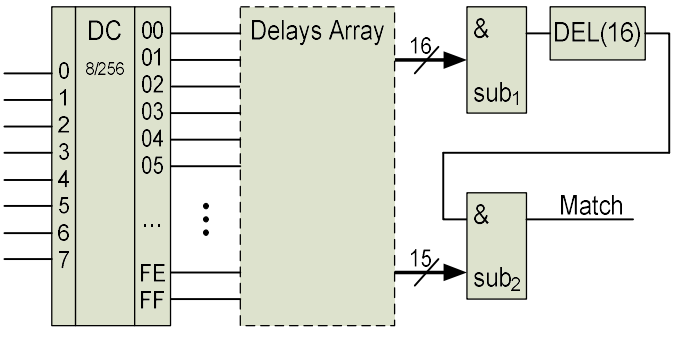


Рис. 4. Схемне рішення DpCAM

На рисунку DEL(1) та DEL(2) – ЦСЗ на один і на два такти відповідно. Таке рішення отримало назву *Схема розпізнавання на базі декодованої асоціативної пам'яті* або схема *DCAM* (Decoded Content-Addressable Memory), тому що повний комплект з ЦК фактично є дешифратором (декодером).

Подальше зменшення складності схеми розпізнавання на базі DCAM можливе шляхом використання техніки часткового розпізнавання, коли довгі патерни розбиваються на коротші фрагменти, що розпізнаються послідовно. При цьому достатньо затримувати лише сигнал часткового збігу замість використання для довгих патернів довгих ЦСЗ на велику кількість тактів. На рис. 4 наведено схему розпізнавання 31-символьного патерну, що побудована за таким принципом. Це рішення отримало назву *Схема розпізнавання на базі частково декодованої АП* або схема *DpCAM* (Decoded partial Content-Addressable Memory).

Більш детальне дослідження свідчить, що використання описаних вище технік зменшення витрат в схемах розпізнавання на ЦК в разі їх паралельного з'єднання (в схему ParCAM) призводить до більш високої економії ресурсів ніж лінійний зріст. Отже, схеми на ЦК мають добру масштабованість за пропускною здатністю.

В розділі також розглянуто використання небайтової обробки даних, а саме, схеми *Hbc, HbcDCAM* та *ParHbcDCAM* на півбайтових компараторах, а також схема *BCAM*, що будується з використанням бінарних діаграм рішень.

Ще зменшити витрати можливо шляхом кластеризації вхідного набору патернів для розбиття схеми на відповідну кількість підсхем меншого розміру

Той факт, що зміст вхідних даних ніяким чином не впливає на характер роботи схеми розпізнавання АП на базі ЦК, реалізує показник передбачуваності пропускної здатності та робить системи, що використовують даний підхід, невразливими до атак на РАСЗТЗІ, що є здійсненням іншого функціонального ПЕ.

Як можна бачити по розглянутих рішеннях, при створенні засобів розпізнавання на ЦК інформація про патерни фактично "прошивається" в апаратну схему, що унеможливлює динамічну реконфігурацію та селективне розпізнавання.

Підсумки та результати формалізації опису властивостей схем розпізнавання на базі АП та ЦК зведені до Таблиці наприкінці розділу.

Підхід на основі геш-функцій та фільтрів Блума.

Фільтр Блума (ФБ) – це абстрактний пристрій, який дозволяє виявити збіг фрагменту заданої послідовності бітів зі зразком із словнику. ФБ складається з двох ключових компонентів (рис. 5): комплекту з K блоків, що обчислюють геш-функції $h_1(x)$, $h_2(x)$, …, $h_K(x)$, та масиву з M бітових комірок (компонент Rg на рисунку). В початковому стані цей регістр бітів (РБ) заповнений нулями.

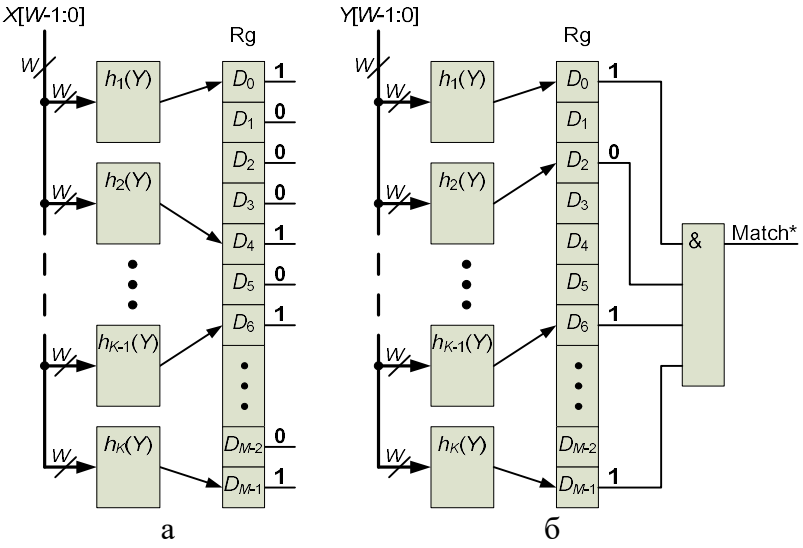


Рис. 5 Принцип дії фільтра Блума: а – програмування; б – розпізнавання

На етапі програмування на входи блоків геш-функцій послідовно подається кожен з n елементів словнику патернів (довжиною W бітів), для цих елементів обчислюються всі K геш-функцій, значення котрих інтерпретуються як адреси комірок у РБ, в які заносяться значення "1".

В процесі функціонування фільтра Блума на його вхід подається фрагмент вхідної послідовності символів (також довжиною W бітів), і також обчислюються значення всіх *K* геш-функцій. По отриманим адресам здійснюється звернення до комірок РБ. Якщо у всіх позиціях, на які вкажуть геш-функції, містяться одиниці, вважається, що вхідна комбінація символів з певною вірогідністю співпадає з одним з патернів, що приймали участь у програмуванні ФБ. Але якщо хоча б одна геш-функція вкаже на комірку з нульовим значенням, це гарантовано свідчить про відсутність збігу. Отже, ФБ функціонує з деякою вірогідністю помилки розпізнавання другого роду (false positive), проте без помилок першого роду (false negative).

ФБ дозволяє економно використовувати ресурси пам’яті (розмір словнику патернів не впливає напряму на розмір РБ: додання нових патернів до вже присутніх призводить лише до підвищення вірогідності помилки розпізнавання, але не до збільшення об’єму потрібної пам’яті). Кількість та складність геш-функцій, отже й апаратурна витратність та продуктивність при даному підході теж не залежать від об’єму словника патернів. Довжина патернів так само не впливає напряму ні на кількість ресурсів пам’яті, ні на продуктивність. Тобто ФБ добре масштабується по двох напрямах: за об’ємом словнику патернів та за довжиною патернів. Дійсно, навіть дуже довгі рядки після перетворення геш-функціями потребують для зберігання ті ж самі *K* комірок РБ.

На жаль, фільтр Блума має важливий недолік, притаманний всім рішенням на базі гешування: розмір вхідної послідовності символів для аналізу має бути фіксованим для обраного набору геш-функції. Тобто, один ФБ здатен розпізнавати патерни тільки однакової довжини. Для подолання цього недоліку будують структуру, що містить кілька ФБ для патернів різної довжини. Це погіршує вартівні ПЕ схем на фільтрах Блума, але не швидкісні показники.

Складності побудови ФБ на ПЛІС пов'язані насамперед с тим фактом, що в базовій схемі фільтра Блума (рис. 5) кільком генераторам геш-функцій потрібно одночасно доступатися до регістру бітів Rg. Якщо реалізувати останній на єдиному запам’ятовуючому пристрої, виникатимуть колізії. Рішення полягає у розбитті РБ на декілька запам'ятовуючих пристроїв залежно від числа геш-функції. Для цього потрібно, по-перше, розподілити виходи блоків реалізації геш-функцій між відповідними ОЗП, по-друге, на функціональність цих блоків накласти додаткові обмеження, щоб їх вихідний діапазон укладався у зменшений об'єм запам'ятовуючого пристрою. Доведено, що такі обмеження підвищують – але не суттєво – вірогідність помилок другого роду фільтра Блума.

В загальному випадку для створення РБ потрібно кілька блоків BRAM. Відповідна схема отримала назву *Повнорозмірний фільтр Блума* або схема *LBF* (Large Bloom Filter). Але в більшості практичних застосувань для побудови МР РАСЗТЗІ можна задіяти окремо так звані схеми міні-ФБ, які є складовими частинами LBF і містять лише по одному блоку BRAM. Назвемо таке рішення *Спрощеним фільтром Блума* або схемою *SBF* (Simplified Bloom Filter).

З метою підвищення швидкодії ФБ можна об’єднувати в паралельну структуру подібно до схеми ParCAM. Але на відміну від ЦК властивості фільтрів Блума не дозволяють досягнути сублінейного закону зростання витрат – апаратні витрати спільної схеми строго пропорційні досягнутому прискоренню, тобто за

пропускною здатністю ФБ масштабується не так добре, як схеми на базі АП/ЦК.

Класична схема фільтра Блума не дозволяє в процесі функціонування вилучати патерни, додані під час програмування. Якщо це потрібно, застосовують *фільтр Блума з лічильниками*, або *схему CBF* (Counting Bloom Filter). Таке рішення дозволяє здійснювати динамічну реконфігурацію без зупинення роботи РАСЗТЗІ.

Необхідність уточнення результатів внаслідок системної помилки ФБ, яке виконується повільніше, ніж розпізнавання, призводить до непередбачуваності пропускної здатності, та робить фільтр Блума вразливим до атак на МСВВ.

Результати формалізації опису властивостей схем розпізнавання на базі ФБ також зведені до Таблиці.

Підхід на основі скінченних автоматів та алгоритму Ахо–Корасік.

Математичний апарат цифрових автоматів (ЦА) успішно використовують для створення обчислювальних систем. Класичний ЦА задається шісткою елементів:

$$A = \{X,\ Y,\ S,\ S_0, f_s, f_y\}, \qquad (1)$$

де X – множина вхідних сигналів, Y – множина вихідних сигналів, S – множина станів автомата, S_0 – початковий стан автомата, f_s – функція переходів з одного стану в інший, f_y – функція виходів автомата.

Але останнім часом в техніці частіше використовують так звані автомати – *розпізнавачі* (classifier). Такий автомат видає на виході активний сигнал тільки у випадку, якщо на його вхід надійшла одна з наперед заданих послідовностей *символів* (комбінацій вхідних сигналів). Коли така ситуація трапляється, розпізнавач переходить в один з так званих *прийнятних* станів. Поки переходи здійснюються між неприйнятними станами, сигнали на виході відсутні. Саме такі автомати використовуються при створенні РАСЗТЗІ.

Відмінність автомата – розпізнавача від класичного ЦА полягає у відсутності множини вихідних сигналів Y та функції виходів f_y, замість яких з'являється множина прийнятних станів F:

$$A = \{X,\ S,\ S_0, f_s,\ F\}. \qquad (2)$$

Тобто автомат – розпізнавач описується не шісткою, а п’ятіркою елементів.

Якщо у виразах (1) та (2) елементи X, Y, S та F є скінченними множинами, ЦА називають *скінченним автоматом* (СА). Це стосується автоматів усіх типів, але останні роки в технічній літературі словосполукою *скінченний автомат* частіше позначають саме автомати – розпізнавачі (в даному дослідженні – також).

Скінченні автомати бувають *детермінованими* (ДСА) та *недетермінованими* (НСА). У детермінованому автоматі на кожному такті можливий лише один перехід лише до одного стану. Як наслідок в кожний окремий момент часу ДСА може перебувати тільки в одному стані. Для НСА вказані обмеження не виконуються.

Алгоритм Ахо–Корасік (АК) є прикладом засобу, який на відміну від багатьох відомих алгоритмів одиночного розпізнавання виявляє у вхідних даних одразу кілька зразків. На етапі побудови алгоритму АК з наданого набору патернів за певними правилами створюється ДСА, який під час функціонування розпізнає потрібні патерни. Назвемо такий ДСА скінченним автоматом Ахо–Корасік (СА-АК).

Теорію та приклади формування СА-АК широко подано в літературі. Однак,

для подальшого дослідження потрібно проаналізувати функцію переходів f_s СА-АК. В роботі виокремлено чотири типи переходів, які присутні в такому автоматі: *прямі* (direct або goto-переходи), *перехресні* (cross), *хибні* (failure) та *післястартові* (restartable). Строго кажучи, останній тип відноситься до перехресних переходів, але внаслідок особливостей технічної реалізації його відділення в окремий тип дозволить спростити поводження з автоматом СА-АК в подальшому дослідженні.

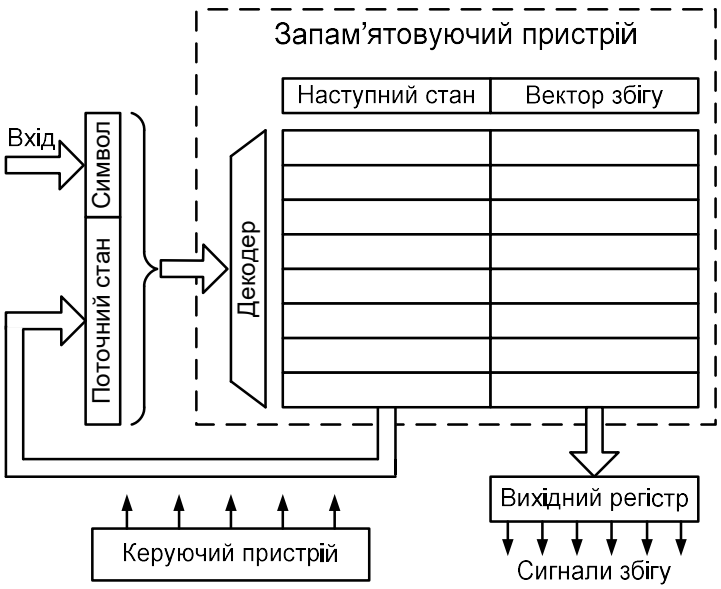


Рис. 6. Узагальнена структурна схема типової реалізації скінченого автомата Ахо–Корасік

Узагальнену структуру апаратної реалізації СА-АК наведено на рис. 6. Його основу складає запам'ятовуючий пристрій (ЗП) в якому зберігається таблиця переходив автомата. Кожна комірка ЗП містить номер наступного стану та вектор збігу. До значення номеру наступного стану, що витягується з пам'яті, шляхом конкатенації додається код символу із вхідній послідовності. Здобуте значення подається на адресний вхід ЗП та обирає відповідний рядок інформації. Вектори збігу (ВЗ) містять одиницю у позиції, що позначає відповідний патерн. Керує роботою автомата керуючий пристрій (КП). Якщо при реалізації схеми використовується зовнішня щодо кристалу ПЛІС пам'ять, називатимемо таке рішення *Базовою схемою СА-АК із зовнішньою пам'яттю* або схемою *ACRAM*. Рішення на базі внутрішньої блокової пам'яті ПЛІС будемо називати *Базовою схемою СА-АК із блоковою пам'яттю* або схемою *ACBRAM*.

Найважливіша перевага рішень на базі СА-АК – незалежність пропускної здатності від об'єму словнику сигнатур та від особливостей патернів, зокрема, від їх довжин, що має наслідком передбачуваність пропускної здатності. Зазвичай СА за кожен такт приймає один символ зі вхідної послідовності. Але на практиці якщо розмір таблиці переходив завеликий, і для її зберігання потрібно використовувати зовнішню відносно ПЛІС пам'ять, кожне звернення до ЗП може займати декілька тактів. Крім того, зовнішня пам'ять повільніша за внутрішню. Отже, зворотна сторона цієї переваги – відносно низька швидкодія, яку до того ж складно нарощувати, що означає погану масштабованість за пропускною здатністю СА-АК.

Основні складнощі, що виникають при реалізації СА-АК на реконфігуровній платформі, пов'язані з організацією ефективного обміну даними із ЗП. Якщо при побудові схеми ACBRAM за рахунок гнучкості конфігурування блоків BRAM існує можливість синтезувати пристрій пам'яті майже довільної структури, то у випадку ACRAM проблема загострюється, тому що бортова пам'ять РУО має фіксовану структуру, не розраховану на ефективну взаємодію з КП.

Як можна бачити, СА-АК потребує незначну кількість ресурсів логіки для створення схеми керування, регістрів та контролера ЗП. Проте об'єм

запам'ятовуючого пристрою може сягати дуже великих значень, тобто підхід характеризується високою ресурсоємністтю щодо пам'яті Збільшення кількості патернів призводить до лавиноподібного зросту ресурсів ЗП, що означає дуже погану масштабованість за об’ємом словнику патернів. Тому переважна більшість досліджень щодо застосування СА-АК в РАСЗТЗІ спрямована саме на зменшення ресурсів пам'яті та стримування їх швидкого зростання.

Серед численних модифікацій СА-АК є рішення, в яких пропонуються різни способи кодування таблиці переходів, також аналізується збіг не тільки префіксів, а й інфіксів патернів. Але в більшості модифікацій зменшення пам'яті досягається шляхом певного поводження з різними типами переходів автомату. В низці робіт було запропоновано та розвинуто техніку застосування конвеєризації при побудові СА-АК: лінійний конвеєр з *H* щаблів дозволяє позбавитися всіх основних перехресних переходів у структурі СА-АК від початкового стану до рівня *H*.

Окрема частка зусиль дослідників була спрямована на покращення не дуже гарних швидкісних ПЕ підходу. Оскільки СА-АК, як будь-який скінченний автомат, обробляє вхідну інформацію строго послідовно – символ за символом, були здійснені спроби прискорити роботу алгоритму АК за рахунок обробки більш, ніж одного символу за такт. Оскільки заздалегідь невідомо, з яким зсувом питомий патерн опиниться у вхідних даних, потрібно організувати паралельну роботу відповідної кількості однакових автоматів (модифікації *ParACRAM* та *ParACBRAM*).

Ще одно рішення, пов’язане з небайтовою обробкою даних, схема *Bit-split*, полягає в заміні автомата, що обробляє 8-бітні символи на декілька однакових паралельно працюючих підавтоматів, які аналізують по 1, по 2 або по 4 біти. За рахунок зменшення "символів", що обробляються, суттєво скорочується розмір алфавіту. Рішення Bit-split є протилежним щодо схем розпізнавання по кілька символів за такт, але воно спрямоване не на прискорення, а на скорочення ресурсів.

Використання зовнішнього ОЗП забезпечує просту реалізацію динамічної реконфігурації схеми СА-АК шляхом перезапису його змісту, тобто повної зміни алгоритму роботи автомата, не зупиняючи процес функціонування РАСЗТЗІ. Щодо показника селективного розпізнавання, він реалізується шляхом формування в пам'яті ЗП кількох таблиць переходів для різних підавтоматів.

Результати формалізації опису властивостей схем розпізнавання на базі СА, що реалізують алгоритм АК, також зведені до Таблиці.

Порівняння підходів.

Як свідчать результати проведеного автором порівняльного аналізу, жоден з досліджених підходів не демонструє явних переваг перед іншими. Кожен має позитивні риси та недоліки та не перевершує конкурентні рішення за всіма ПЕ. Тому виникає потреба в методах, які дозволили би поєднати різні підходи в єдиному пристрої, максимізуючи ефективність за рахунок реалізації їх переваг.

Для створення таких методів необхідно розробити інструмент кількісної оцінки та порівняння різнорідних технічних рішень в єдиному метричному просторі та скласти відповідні розрахунки для кожного з напрямів.

У четвертому розділі автором розроблено методи, які дозволяють кількісно оцінювати і порівнювати схеми розпізнавання та їх складові, що реалізовані реконфігуровними засобами на базі різних технологій та підходів до побудови МР

РАСЗТЗІ. Для цього виявлено параметри ефективності, що можуть бути оцінені кількісно, та технічні характеристики компонентів РАСЗТЗІ, що їм відповідають.

Таблиця

Порівняння основних підходів до побудови МР РАСЗТЗІ

№	Показник ефективності		Підхід		
			Асоціативна пам'ять	Фільтр Блума	Скінчен. автомат Ахо–Корасік
1.	Витрати логіки		---	+	+++
2.	Витрати пам'яті	розподіленої	---	+	+++
3.		блочної	+++	+	---
4.		зовнішньої	+++	+++	---
5.	Швидкодія		+++	+	-
6.	Передбачуваність пропускної здатності		+++	---	+++
7.	Функціональні показники	здатність протидіяти атакам на РАСЗТЗІ	+++	---	+
8.		динамічна реконфігурація	---	+	+++
9.		селективне розпізнавання	---	+	+++
10.		режим запобігання вторгнень (для МСВВ)	+++	-	---
11.	Масшбо-ваність	за пропускною здатністю	+	-	-
12.		за об'ємом словнику	-	+++	---
13.		за довжиною патернів	-	+++	+++
14.	Використання надлишковості		+++	---	+
15.	Суттєвий недолік, який зводить нанівець головні переваги підходу		Завелике споживання ресурсів	Фіксована довжина патернів	Вибухоподібний зріст об'єму пам'яті

Позначення: "+++" – суттєва перевага, "+" – помірна перевага, "---" – суттєвий недолік, "-" – помірний недолік.

Оскільки властивості словнику патернів, що входять до складу бази даних сигнатур РАСЗТЗІ, мають важливе значення при реалізації методів комбінування, в розділі введені потрібні визначення та розроблені техніки поводження з патернами.

Подамо множину патернів, що мають розпізнаватися, у наступному вигляді:

$$P = \{p_1, p_2, p_3, \dots p_k, \dots p_\sigma / \sigma, \Omega, m_{\min}, m_{\max}, \delta, \mu, \mu_z, \nu\}, \quad (3)$$

де $p_1, p_2, p_3, \dots p_k, \dots p_\sigma$, – власне набір патернів, σ – потужність множини, Ω – загальна кількість символів, $m_{\min}$ – довжина найкоротшого патерну в наборі, $m_{\max}$ – довжина найдовшого патерну в наборі, δ – функція розподілу довжин, μ – перша функція самоподоби, μ_z – перша часткова функція самоподоби, ν – друга функція самоподоби. Патерни p_k є фіксованими послідовностями символів, код кожного з котрих належить до певного алфавіту Σ. У випадку байтового кодування $\Sigma = \{00_{16}, 01_{16}, 02_{16}, \dots, FF_{16}\}$.

В якості основи подальших розрахунків автором запропоновано наступну техніку впорядкування патернів у наборі. Відсортуємо всі патерни в множині P за

зростанням довжини, як наведено на рис. 7. Складені з квадратів стовпчики тут зображають складені з символів рядки.

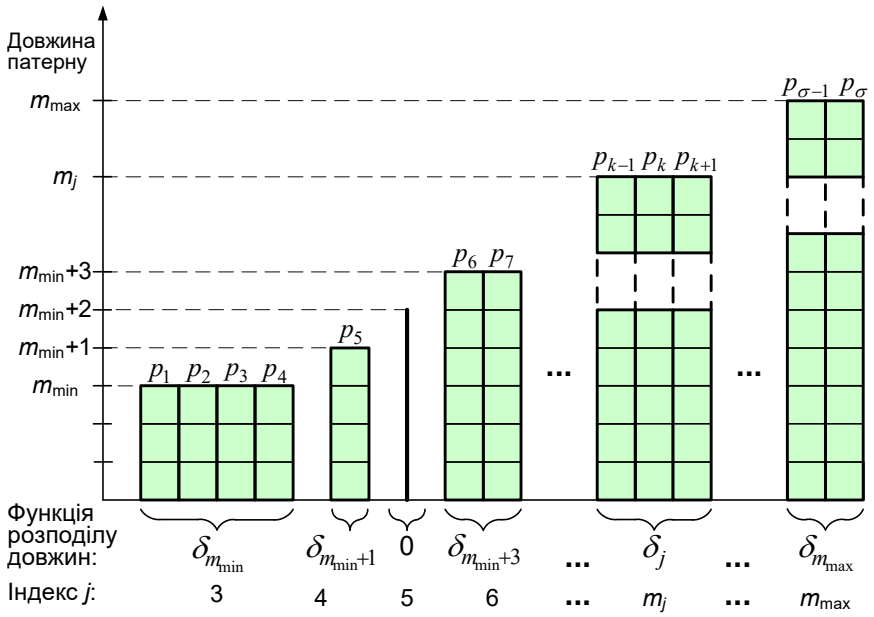


Рис. 7. Техніка впорядкування пакетами патернів однакової довжини

Назвемо *пакетом* сукупність патернів однакової довжини, не звертаючи уваги, як само впорядковані патерни всередині пакету. Введемо індекс j таким, що співпадає з довжиною патернів в пакеті $j = m_{\min}, m_{\min}+1, m_{\min}+2, \ldots m_j, \ldots m_{\max}$, де $m_{\min}$ – довжина найкоротшого патерну, $m_{\max}$ – довжина найдовшого патерну, причому кожне наступне значення цього індексу обов'язково на одиницю більше за попереднє, тобто в нумерації немає пропусків. Тоді довжина кожного патерну в наборі співпадатиме з індексом його пакету: $m_j = j$. Зворотне твердження хибне, тому що для деяких індексів j патерни відповідної довжини можуть бути відсутні.

Функцію розподілу довжин δ як залежність від індексу j визначимо рівною кількості патернів у відповідному пакеті: $\delta(j) = \delta_j$. В прикладі на рис. 7 $j = 3, 4, 5, \ldots m_{\max}$, $\delta(3)=4$, $\delta(4)=1$, $\delta(5)=0$, $\delta(6)=2$, $\delta(m_{\max})=2$.

За допомогою функції розподілу довжин зручно обчислити кількість ненульових пакетів ξ та загальну кількість символів Ω в наборі. Функції самоподоби, які також розглянуті в даному розділі, дозволяють в різний спосіб кількісно оцінювати надмірність словнику патернів.

На основі аналізу сформульованих у попередньому розділі ПЕ було з'ясовано, що для кількісної оцінки та порівняння різних технічних рішень щодо МР РАСЗТЗІ достатньо використовувати лише ресурсні та часові показники. Але якщо часові характеристики нескладно звести до єдиної одиниці виміру, то для показнику ресурсних витрат ситуація є складнішою. Щоб забезпечити методичну строгість порівняння різних технічних рішень, необхідно звести підрахунок ресурсів різного типу до якоїсь єдиної умовної одиниці. В якості такої одиниці автором запропоновано використовувати логічну таблицю (LUT) як мінімальний елемент структури ПЛІС, тобто здійснювати підрахунки в *умовних логічних таблицях* (УЛТ).

Використовуючи такий підхід, значення R ресурсів, що потрібні для синтезу деякого компоненту обчислювальної структури, може бути подано у вигляді:

$$R = L + \alpha F + \beta B + \gamma M \,, \tag{4}$$

де L – об'єм ресурсів логіки ПЛІС (кількість ЛТ), F – ресурси розподіленої пам'яті ПЛІС (кількість тригерів), B – об'єм ресурсів блокової пам'яті ПЛІС (число блоків BRAM), M – об'єм ресурсів зовнішньої пам'яті – бортової пам'яті РУО (у Мбайтах), α, β, γ – коефіцієнти нормалізації відповідних ресурсів відносно УЛТ.

Ресурсні та часові характеристики реконфігуровних компонентів нескладно знайти шляхом синтезу їх цифрових схем за допомогою інструментальних засобів (фірмових пакетів) створення конфігурацій для ПЛІС. Але цей процес потребує багато часу, що унеможливлює його використання в циклі процедури оптимізації при використанні методів створення ефективних РАСЗТЗІ.

Тому автором було запропоновано ***Метод прискореного обчислення технічних характеристик*** реконфігуровних цифрових пристроїв без виконання процедури їх синтезу. Суть методу полягає у створенні для цифрової схеми розпізнавання так званої функції оцінки (ФО) θ. Така функція, маючи на вході заданий набір патернів та параметрів РУО, здатна з певною точністю обчислити та отримати на виході в якості результату значення об'єму ресурсів R (в УЛТ), що вживатиме схема, яка буде розпізнавати цей набір патернів, та значення часової затримки сигналу в неї T (в одиницях часу), яку вона матиме після синтезу:

$$\theta = \{ R, T \} = \left\{ \theta_R \left(P, \Theta, \Psi \right), \theta_T \left(P, \Theta, \Psi \right) \right\}, \tag{5}$$

де P – набір патернів, що потрібно розпізнавати, Θ – сукупність характеристик прискорювача, Ψ – сукупність вимог користувача системи захисту.

Функції розрахунку ресурсів θ_R та часу θ_T, тобто *ресурсна складова* (РС) та *часова складова* (ЧС) відповідно, в загальному випадку також залежать від набору патернів, характеристик РУО, та вимог користувача, які є кількісними обмеженнями та параметрами задіяного підходу до побудови схеми розпізнавання.

Набір патернів P визначається виразом (3). Властивості РУО задаються множиною його характеристик, з яких більшість відноситься до ПЛІС:

$$\Theta = \{ x, y, z, p, U, B_{\text{FPGA}}, L_{\text{FPGA}}, F_{\text{FPGA}}, T_{\text{FPGA}}, M_{\text{RA}}, \alpha, \beta, \gamma \}, \tag{6}$$

де x – число входів ЛТ мікросхеми ПЛІС, що задіяна в РУО, y – здатність навантаження (fan-outmax) логічних елементів ПЛІС, z – максимальне значення затримки ЦСЗ, створюваної на бази ЛТ, p – кількість портів блокової пам'яті ПЛІС, U – розмір блоків пам'яті BRAM в мікросхемі ПЛІС (у Кбітах), BFPGA – кількість блоків BRAM в мікросхемі ПЛІС, LFPGA – кількість ЛТ в ПЛІС, FFPGA – кількість тригерів в ПЛІС, TFPGA – максимально можлива швидкодія ПЛІС (період тактової частоти), MRA – об'єм бортової пам'яті РУО (у Мбайтах), α, β, γ – згадані вище коефіцієнти зведення ресурсів різного типу до УЛТ, які в загальному випадку залежать від типу використаної ПЛІС.

До множини Ψ належать кількісні обмеження, що накладає користувач системи захисту, яки мають сенс порогових значень для певних технічних характеристик та параметри підходу до побудови РАСЗТЗІ (наприклад, фактор

хибного розпізнавання для схем на базі фільтра Блума).

Після обчислення ФО для всіх компонентів МР, можна знайти об'єм ресурсів та часову затримку для всього пристрою.

Далі в розділі розглянуто приклади побудови ФО для базових схем обраних підходів та їх найбільш ефективних модифікацій, що розглянуті в попередньому розділі. Отримані залежності дозволяють, з одного боку, експериментально перевірити теоретичні положення щодо запропонованих методів комбінування, з іншого – можуть бути використані розробниками РАСЗТЗІ на практиці.

З великої кількості модифікацій підходів до побудови МР РАСЗТЗІ та технік підвищення їх ефективності є сенс обмежитися розгляданням процесів створення ФО лише для найбільш ефективних варіантів побудови блоків розпізнавання. Згідно проведеному в попередньому розділі дослідженню до таких варіантів відносяться: базова схема на АП BsCAM, модифікована схема на АП DCAM, модифікована схема на АП DpCAM, повнорозмірний фільтр Блума LBF, спрощений фільтр Блума SBF, базова схема СА-АК із зовнішньою пам'яттю ACRAM, базова схема СА-АК із блоковою пам'яттю ACBRAM, схема СА-АК за конвеєрною схемою PipACBRAM, а також двокаскадна схема HRCmp.

В зв’язку з простотою та регулярністю структури схем розпізнавання на ЦК ФО для блоків розпізнавання на базі АП може бути заснована на прямому підрахунку потрібних ресурсів та часових затримок. В деяких випадках результати можуть бути уточнені внаслідок використання певних технічних особливостей реалізації задіяних підходів. Враховуючи це, в роботі було запропоновано *Метод створення ФО для схем АП на базі ЦК*, який в загальному випадку складається з етапу прямого підрахунку та декількох етапів уточнення отриманих співвідношень.

Після виконання потрібних дій, докладно описаних в розділі, згідно структури базової схеми на компараторах (рис. 2) та співвідношення (4) отримуємо (у другому теоретичному наближенні) для РС ФО схеми BsCAM наступний вираз:

$$R^{**}_{\text{BSCAM}} = \sum_{j=m_{\min}}^{m_{\max}} \delta_j \left(\Lambda(x) j + \left\lceil \frac{j-1}{x-1} \right\rceil \right) +$$

$$+ \alpha \left(8 m_{\max} + m_{\min} \left\lceil \frac{\sigma - 1}{y-1} \right\rceil + \sum_{j=m_{\min}+1}^{m_{\max}} \left\lceil \left(\sum_{i=j}^{m_{\max}} \delta_i - 1 \right) \Big/ y - 1 \right\rceil \right), \qquad (7)$$

де $\Lambda(x)$ – функція-кваліфікатор, яка введена для врахування числа входів логічних таблиць ПЛІС та визначена наступним чином:

$$\Lambda(x) = \begin{cases} 1, & x \geq 8 \\ 2, & x < 8 \end{cases}.$$

Вираз (7), який визначає РС ФО для схеми для базової схеми на цифрових компараторах BsCAM залежить від наступних параметрів словнику патернів та ПЛІС: $\theta_R = \theta_R(P, \Theta) = R^{**}_{\text{BSCAM}}(\sigma, m_{\min}, m_{\max}, \delta, x, y, \alpha)$. Зауважимо, що в даному виразі відсутні функції самоподоби множини патернів. Тобто базова схема на АП

BsCAM не використовує надмірність, яка присутня в базі даних сигнатур. Наявність лише одного з коефіцієнтів нормалізації (α) говорить про те, що при реалізації на ПЛІС схеми BsCAM використовується два види ресурсів (ЛТ та тригери).

ЧС ФО схеми BsCAM в другому емпіричному наближенні приймає наступний вигляд:

$$T^{**} = S/\left(A_T + B_T R + C_T R^2\right), \tag{8}$$

де A_T, B_T та C_T – заздалегідь знайдені емпіричним шляхом коефіцієнти квадратичної апроксимації функціональної залежності максимальної тактової частоти синтезованої в ПЛІС цифрової схеми від ресурсних витрат, S – коефіцієнт ділення частоти, який вказує, в скільки разів відрізняється тактова частота всередині "швидких" та "повільних" зон згідно техніки поділу цифрової схеми на часові зони.

РС ФО для модифікації DCAM на компараторах, як випливає з розрахунків, у другому теоретичному наближенні має наступний вигляд:

$$R^{**}_{\mathrm{DCAM}} = 256\Lambda(x) + \sum_{j=2}^{m_{\max}}\left(\left\lceil\frac{j-1}{z+1}\right\rceil\sum_{s=0}^{255}\mathrm{NotZ}(\mu(s,j))\right) +$$

$$+\sum_{j=m_{\min}}^{m_{\max}}\delta_j\left\lceil\frac{j-1}{x-1}\right\rceil + \alpha\sum_{s=0}^{255}\left\lceil\left(\sum_{j=2}^{m_{\max}}\mathrm{NotZ}(\mu(s,j)) + \mu(s,j) - 2\right)\Big/ y-1\right\rceil, \tag{9}$$

де NotZ – функція нерівності нулю, $\mu(s,j)$ – перша функція самоподоби, яка чисельно дорівнює сумарної кількості символів з кодом s, розташованих на j-ої позиції всіх патернів словнику.

Згідно (9) РС ФО для схеми DCAM залежить від наступних параметрів словнику патернів та ПЛІС: $\theta_R = R^{**}_{\mathrm{DCAM}}(m_{\min}, m_{\max}, \delta, \mu, x, y, z, \alpha)$. Наявність в даному виразі функції самоподоби свідчить, що дана модифікація підходу на АП використовує надмірність бази даних сигнатур.

ЧС ФО для схеми DCAM така ж сама, що й для базової схеми (8).

Вираз для РС ФО модифікації DpCAM виявляється ще складнішим:

$$R^{**}_{\mathrm{DpCAM}} = 256\,\Lambda(x) + \sum_{j=2}^{z-1}\left(\sum_{s=0}^{255}\left\lceil\frac{j-1}{z+1}\right\rceil\mathrm{NotZ}(\mu_{z-1}(s,j)) + \left\lceil\frac{\mu_{z-1}(s,j)-1}{y-1}\right\rceil\right) +$$

$$+\sum_{j=m_{\min}}^{m_{\max}}\delta_j\left(\left\lfloor\frac{j}{z-1}\right\rfloor\left(\left\lceil\frac{z}{x}\right\rceil+1\right) + \left\lceil\frac{j \bmod (z-1)}{x}\right\rceil+1\right) +$$

$$+\sum_{s=0}^{255}\left\lceil\left(\sum_{j=2}^{z-1}\mathrm{NotZ}(\mu_{z-1}(s,j)) - 1\right)\Big/(y-1)\right\rceil, \tag{10}$$

де $\mu_{z\text{-}1}(s,j)$ – перша часткова функція самоподоби, яка чисельно дорівнює сумарної кількості символів з кодом s, розташованих періодично на позиціях

$k(z-1)(j-1)$, де $j = 2, 3, \ldots (z-2)$, $k = 1, 2, \ldots \lceil m_j/(z-1) \rceil$ всіх патернів словнику.

Згідно (10) РС ФО для схеми DCAM подібна до схеми DCAM: $\theta_R = R^{**}_{\mathrm{DCAM}}(m_{\min}, m_{\max}, \delta, \mu_{z-1}, x, y, z, \alpha)$. Наявність у даному виразі однієї з функцій самоподоби свідчить про використання схемою надмірності бази сигнатур.

ЧС ФО для схеми DpCAM така сама, що й для базової схеми (8).

Реалізація на ПЛІС підходу на базі фільтра Блума призводить до побудови більш складної та менш регулярної структури. Логічні та тригерні ресурси використовуються тут для створення таких складових ФБ, як генератор геш-функцій (ГГФ) та регістр бітів – РБ. Створення РБ потребує також використання ресурсів блокової пам'яті BRAM. Всередині структури схеми розпізнавання масово присутні допоміжні та керуючі підсхеми, що ускладнює вирази. Внаслідок цих причин вираз для РС ФО повнорозмірного фільтра Блума LBF виявляється більш складнім:

$$
\begin{aligned}
R_{\mathrm{LBF}} = {} & E \log_2 U \left(\left\lceil \frac{8j}{x} \right\rceil + \left\lceil \frac{\left\lceil \frac{8j}{x} \right\rceil - 1}{x-1} \right\rceil \right) + \\
& + \frac{E \log_2 U + 1}{\left(x - \left\lceil \log_2 \left\lceil \frac{\delta_j E}{U \ln 2} \right\rceil \right\rceil \right) / 2} + \left\lceil \log_2 \left\lceil \frac{\delta_j E}{U \ln 2} \right\rceil \right\rceil \left(\left\lceil \frac{8j}{x} \right\rceil + \left\lceil \frac{\left\lceil \frac{8j}{x} \right\rceil - 1}{x-1} \right\rceil \right) + \\
& + \left\lceil \frac{\delta_j E}{U \ln 2} \right\rceil \left(\left\lceil \frac{\left\lceil \frac{E}{p} \right\rceil - 1}{x-1} \right\rceil + \left\lceil \frac{E}{p} \right\rceil \left(\left\lceil \frac{\log_2 U}{\left\lceil \frac{x-1}{2} \right\rceil} \right\rceil + 4 \right) \right) + \alpha \left(E \log_2 U \left(\left\lceil \frac{\left\lceil \frac{8j}{x} \right\rceil - 1}{x-1} \right\rceil - 1 \right) + \right. \\
& \left. + \left\lceil \log_2 \left\lceil \frac{\delta_j E}{U \ln 2} \right\rceil \right\rceil \left(\left\lceil \frac{\left\lceil \frac{8j}{x} \right\rceil - 1}{x-1} \right\rceil - 1 + \left\lceil \frac{\delta_j E}{U \ln 2} \right\rceil \left\lceil \frac{E}{p} \right\rceil \log_2 U \right) + \beta \left\lceil \frac{\delta_j E}{U \ln 2} \right\rceil \left\lceil \frac{E}{p} \right\rceil, \right. \qquad (11)
\end{aligned}
$$

де E – фактор хибного розпізнавання, що чисельно дорівнює ступеню двійки, обернена величина до якої є дозволеною вірогідністю помилки розпізнавання $\rho_{дозв.}$, яку не має перевищувати Фільтр Блума: $E = -\lfloor \log_2 \rho_{дозв.} \rfloor$. В той же час параметр E дорівнює числу K геш-функцій у фільтрі Блума.

Присутність у виразі (11) індексу j нагадує, що ця формула дає значення використаних ресурсів лише для одного ФБ, який здатен розпізнавати патерни довжиною в j символів. Для підрахунку витрат на повну схему розпізнавання на базі ФБ потрібно підсумувати ресурси на створення кожного ФБ для всіх ненульових довжин патернів.

РС ФО для спрощеної схеми фільтра Блума SBF має простіший вигляд:

$$R_{\mathrm{SBF}} = E\left\lceil \log_2 \frac{E\cdot\delta_j}{\ln 2}\right\rceil\left(\left\lceil\frac{8j}{x}\right\rceil+(\alpha+1)\left\lceil\frac{\left\lceil\frac{8j}{x}\right\rceil-1}{x-1}\right\rceil-\alpha\right)+$$

$$+\left\lceil\frac{E}{p}\right\rceil\left(\alpha\left\lceil \log_2 \frac{E\cdot\delta_j}{\ln 2}\right\rceil+\beta+\left\lceil\frac{\left\lceil \log_2 \frac{E\cdot\delta_j}{\ln 2}\right\rceil}{\left\lfloor\frac{x-1}{2}\right\rfloor}\right\rceil+4\right)+\left\lceil\frac{\left\lceil\frac{E}{p}\right\rceil-1}{x-1}\right\rceil. \qquad (12)$$

Присутність в виразах (11) та (12) коефіцієнтів нормалізації α та β свідчить, що схеми на базі фільтрів Блума вживають ресурси трьох типів: логічні таблиці, тригери та блокову пам'ять.

При складанні ЧС ФО для схем на базі ФБ можна також використовувати вираз (8), оскільки затримку доступу до блокової пам'яті шляхом конвеєризації можна зменшити до одного такту. Але потреба додаткового етапу уточнення результатів роботи ФБ (в зв'язку з системною помилкою розпізнавання другого роду) за певних обставин може зменшити швидкодію пристрою в декілька разів.

При створенні МР РАСЗТЗІ на базі СА-АК потрібні логічні ресурси та тригери для побудови керуючого пристрою – КП, контролера зовнішнього оперативного запам’ятовуючого пристрою (КОЗП), якщо ОЗП зовнішній, а ПЛІС не містить апаратно реалізованого контролера, а також ресурси пам’яті.

Як свідчить аналіз, кількість ресурсів, що потрібні для побудови КП та КОЗП незначною мірою залежить від словнику патернів. Фактично ця залежність зводиться лише до зміні розрядності регістрів КП та КОЗП при зміні розміру ЗП скінченого автомату. Запропонована *Техніка підрахунку ресурсів керуючих пристроїв для схем СА-АК* полягає в тому, що кількості ЛТ та тригерів, потрібних для створення кожного з керуючих пристроїв, подаються як функціональна залежність від розрядності шини адреси пам'яті. Експериментально було з'ясовано, що лінійна апроксимація забезпечує прийнятну точність для ресурсів обох типів. Відповідні коефіцієнти (окремо для ЛТ та тригерів) знаходяться емпірично в результаті виконання синтезу декількох тестових варіантів пристроїв.

Кількість ресурсів пам'яті, потрібних для створення скінченного автомату АК, складно подати у вигляді аналітичної функції від змінних, що є властивостями словнику патернів, та параметрів, що є властивостями РУО. Суть запропонованої автором *Техніки підрахунку ресурсів пам'яті для схем СА-АК* полягає у використанні реалізованого програмними засобами функціонального перетворення ProcAC(P, Trdr, $\mathrm{Tr_{cr}}$, $\mathrm{Tr_{fl}}$, $\mathrm{Tr_{rs}}$), яке для обраного варіанту реалізації СА-АК з поданого на його вхід набору патернів P шляхом швидкого будування в пам'яті комп'ютера віртуального скінченного автомата Ахо–Корасік знаходить та подає на вихід кількості прямих (Trdr), перехресних ($\mathrm{Tr_{cr}}$), хибних ($\mathrm{Tr_{fl}}$) та післястартових ($\mathrm{Tr_{rs}}$) переходів. З отриманих кількостей переходів згідно з використаним кодуванням та задіяною технікою скорочення запам’ятовуючих ресурсів розраховуються об’єми

пам'яті, що потрібні для зберігання переходів кожного типу та загальний об’єм ресурсів ЗП:

$$B_{\text{ЗП}} = f_{\text{dr}}(\text{Tr}_{\text{dr}}) + f_{\text{cr}}(\text{Tr}_{\text{cr}}) + f_{\text{fl}}(\text{Tr}_{\text{fl}}) + f_{\text{rs}}(\text{Tr}_{\text{rs}}) + f_{\text{MV}},$$

де $f$dr(), $f$cr(), $f$fl(), $f$rs() та f_{MV} – функціонали, які визначають, скільки ресурсів пам'яті (в байтах) потрібно для зберігання переходів кожного типу та векторів збігу відповідно. Для більшості реалізації СА-АК дані функціонали являють собою множення на коефіцієнт, який дорівнює кількості байтів, потрібних для зберігання даних про один перехід відповідного типу.

Використовуючи запропоновані методи, отримуємо наступне представлення РС ФО для схеми СА-АК із зовнішньою пам'яттю ACRAM:

$$\theta^R_{\text{ACRAM}} = A_{\text{ЛКП}} + A_{\text{ЛКОЗП}} + (B_{\text{ЛКП}} + B_{\text{ЛКОЗП}})(\lceil \log_2 M_{\text{РУО}} \rceil + 20) +$$

$$+ \alpha\,(A_{\text{ТКП}} + + A_{\text{ТКОЗП}} + (B_{\text{ТКП}} + B_{\text{ТКОЗП}})(\lceil \log_2 \rceil + 20)) + \gamma B_{\text{ЗП}}, \qquad (13)$$

де $A_{\text{ЛКП}}$, $B_{\text{ЛКП}}$, $A_{\text{ТКП}}$, $B_{\text{ТКП}}$, $A_{\text{ЛКОЗП}}$, $B_{\text{ЛКОЗП}}$, $A_{\text{ТКОЗП}}$, $B_{\text{ТКОЗП}}$ – коефіцієнти лінійної апроксимації, знайдені емпіричним шляхом для схем КП та КОЗП – окремо для витрат логіки (з літерою "Л" в індексах) та тригерів (з літерою "Т" в індексах).

Коефіцієнти нормалізації α та γ у виразі (13) свідчать, що схема ACRAM використовує три типи ресурсів: логічні таблиці, тригери та зовнішню пам'ять.

ЧС ФО для схеми ACRAM знаходиться як:

$$\theta^T_{\text{ACRAM}} = T_{\text{ACRAM}} = T_{\text{RAM}} \frac{N_{\text{dr}} R_{\text{dr}} + N_{\text{cr}} R_{\text{cr}} + N_{\text{fl}} R_{\text{fl}} + N_{\text{rs}} R_{\text{rs}} + N_{\text{MV}} R_{\text{MV}}}{5 R_{\text{ОЗП}}}, \qquad (14)$$

де Ndr, N_{cr}, N_{fl}, N_{rs} та N_{MV} – кількості звернень до ОЗП за такт роботи СА у випадку здійснення прямого, перехресного, хибного, післястартового переходу та отримання коду прийнятного стану відповідно, Rdr, R_{cr}, R_{fl}, R_{rs} та R_{MV} – об’єми пам'яті (в байтах), потрібні для зберігання відповідно прямих, перехресних, хибних і післястартових переходів та векторів збігу відповідно, RОЗП – сумарний об’єм пам’яті, потрібної для всіх переходів автомату та векторів збігу.

При реалізації СА-АК на базі пам'яті BRAM виникають ті ж складності, що і при побудові фільтрів Блума. Тому вирази для РС і ЧС ФО схеми АК із внутрішньою пам'яттю ACBRAM (наведені в роботі) виявляються більш складними.

Підкреслимо, що здобуті автором вирази (7) – (14) та інші, наведені в дисертаційної роботі, є функціональними залежностями кількісних характеристик відповідних схем розпізнавання від параметрів реконфігуровного обчислювача (як констант) та характеристик заданого набору патернів (як змінних).

У п'ятому розділі автором сформульовано, вдосконалено та розвинуто методи, що базуються на ідеї комбінування (сумісного використання) різних підходів до побудови МР РАСЗТЗІ для максимізації його ефективності за рахунок використання переваг кожного з підходів. В якості цільової функції оптимізації використовується певний технічний параметр МР, який залежить від відповідних параметрів всіх компонентів, з яких він складається. До основних параметрів

належать використані апаратні ресурси та часові властивості. За вимогами користувачів можуть бути задіяні також інші параметри, похідні від основних.

Ресурсні параметри. Ресурси R, потрібні для синтезу будь-якої цифрової схеми, складаються з ресурсів R_i, потрібних для створення кожного i-го компоненту.

Часові параметри. Кожен компонент обробляє дані з певною швидкодією, яка визначається значенням T часу розповсюдження сигналу від входу до виходу. В разі паралельного з'єднання компонентів, кожен з котрих має свій параметр T_i, час затримки цифрової схеми в цілому дорівнює затримці найповільнішого з компонентів. При послідовному з'єднанні компонентів час затримки всього модулю знаходиться як сума затримок кожного з них.

Пропускна здатність C модулю розпізнавання визначається як кількість інформації, що обробляється за одиницю часу. Цей параметр формується з аналогічних параметрів C_i кожного з компонентів, кожен з яких знаходиться шляхом ділення величини M_i на T_i, де M_i – об'єм інформації, що обробляється i-м компонентом. В разі паралельного з'єднання пропускна здатність компонентів додається. У випадку послідовного з'єднання компонентів пропускна здатність всього МР визначається найменш продуктивним з них.

Похідні параметри. З метою більш ретельного порівняння можливих технічних рішень дослідники та розробники РАСЗТЗІ використовують більш складні параметри, які походять від основних і також можуть бути задіяні в якості цільової функції при створенні критеріїв оптимізації. Прикладом такого параметру є ефективність пропускної здатності Q, яка має сенс відношення пропускної здатності до витрачених ресурсів: Q = C / R, або приведена ефективність пропускної здатності E, що знаходиться як добуток пропускної здатності на кількість символів у словнику патернів Ω, поділений на значення ресурсів: $E = (\Omega C) / R$. В разі потреби можуть бути використані й більш складні та витончені похідні технічні параметри.

Критерії оптимізації. Критерієм оптимізації при застосуванні методів підвищення ефективності РАСЗТЗІ завжди є досягнення мінімуму або максимуму цільової функції, в якості якої виступає чисельне значення будь-якого з розглянутих вище технічних параметрів МР, яке відбувається при виконанні наданих умов та накладених обмежень:

$$K \to \max \Big/ Constr\text{'}$$

де K – узагальнене позначення будь-якого критерію оптимізації,

Constr – обмеження, які є сукупністю наступних чинників:

$$Constr = \{\Theta, \Psi, \chi\},$$

де Θ – сукупність характеристик реконфігуровного прискорювача, які розшифровані в (6), Ψ – згадана в (5) сукупність кількісних обмежень та параметрів підходів, χ – множина опцій, що задає користувач, в якості яких використовуються якісні показники ефективності РАСЗТЗІ, такі як можливість динамічної реконфігурації Din, опція протидії спрямованим на МСВВ атакам $AtDef$, спроможність МСВВ функціонувати в режимі запобігання вторгнень Ips, опція незалежності швидкодії від вхідних даних $Tstab$ тощо:

$$\chi = \{Din, AtDef, Ips, Tstab, \ldots\}.$$

Якщо заданий критерій, що має сенс швидкодії, пропускної здатності або походить від них, застосовуються обмеження по ресурсах виду: $R < R_{\text{FPGA}}$ $(1 - \Delta/100)$ або $R < R_{\text{FPGA}} - \Delta\,\text{R}$, де R_{FPGA} – об'єм ресурсів ПЛІС, Δ – порогове значення (у відсотках) невикористаної частки ресурсів ПЛІС введення якої обумовлено тим фактом, що наближення розміру цифрової схеми, що синтезуються, до максимуму ресурсів призводить до вибухоподібному зросту складності та витрат часу на синтез, Δ_{R} – порогове значення (в абсолютних одиницях) невикористаної частки ресурсів ПЛІС.

В разі використання ресурсного критерію оптимізації, тобто коли необхідно мінімізувати витрати, обмеженнями є вимоги щодо швидкодії, які задаються у наступному вигляді: $T_{\text{FPGA}} < T < T_{\text{USR}}$, де T_{FPGA} – порогове значення показника швидкодії ПЛІС (мінімально можлива затримка) як характеристика РУО, T_{USR} – порогове значення швидкодії, задане користувачем.

Суть ***Методу паралельного комбінування*** (МПрКм), сформульованого автором, полягає в поділі набору патернів, які повинні відшукуватися модулем розпізнавання, на підгрупи та синтезі в складі модулю розпізнавання такої ж кількості блоків розпізнавання (БР), кожен з яких, залежно від використаного підходу, найбільш результативно відшукує патерни відповідної підгрупи. Найвища ефективність при цьому досягається за рахунок охоплення обох процесів – поділу патернів та вибору комбінації БР – загальною процедурою оптимізації.

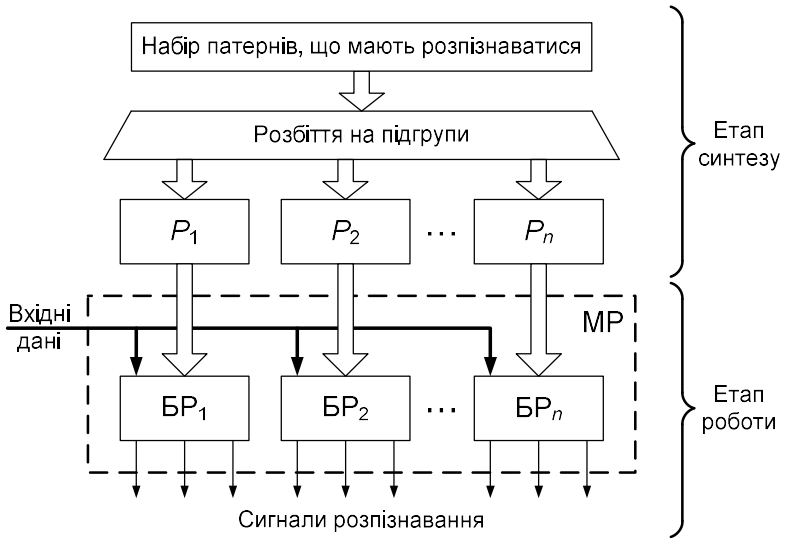


Рис. 8. Схематичне подання структури методу паралельного комбінування

На рис. 8 наведено схематичне зображення структури МПрКм. На етапі синтезу набір патернів P, що підлягає розпізнаванню, розбивається на n підгруп P_i, $i = 1, 2, \ldots n$. Після синтезу в процесі функціонування кожний блок розпізнавання БР_i в складі МР здійснює розпізнавання патернів відповідної підгрупи в потоці вхідних даних, що подаються одночасно на входи всіх БР. В разі виявлення збігу фрагменту вхідних даних з якимось з патернів активується відповідний сигнал розпізнавання.

Змінними параметрами в процесі виконання процедури оптимізації є, з одного боку, варіанти поділу патернів – кількість підгруп n та склад P_i кожної з них, з іншого – комбінація БР_i, які вибираються з бібліотеки готових компонентів, створених з використанням підходів та модифікацій, формалізованих у третьому розділі.

Слід зауважити, що при об'єднанні методом МПрКм різних за своєю природою підходів до побудови схем розпізнавання в загальному випадку виникає

потреба в узгодженні швидкісних характеристик БР$_i$. Для цього, наприклад, застосовується прискорення повільнішої технології шляхом розпаралелювання.

Недоліком розглянутого вище методу МПрКм є вимога того, щоб кожен з патернів був розпізнаним повністю, інакше відповідний сигнал тривоги не активується. Але щоб зробити висновок, що поточний фрагмент вхідних даних не збігається з патерном, досить виявити розбіжність всього в одному символі. Тоді порівняння решти рядку стане зайвим. ***Метод послідовного каскадування*** (МПсКс), сформульований автором, використовує цей факт для підвищення ефективності процесу розпізнавання. Його суть полягає у розбитті операції порівняння з патерном на послідовні етапи, на кожному з котрих здійснюється часткове розпізнавання відповідного фрагменту за умови виявлення збігу на попередньому етапі.

Внаслідок особливостей методу МПсКс до характеристик різних каскадів висуваються різні вимоги, що створює сприятливі умови для використання в них різних за принципами побудови компонентів з метою підвищення ефективності рішення в цілому.

На рис. 9 наведено схематичне зображення структури МПсКс. На етапі синтезу кожний патерн всередині вхідного набору P, ділиться на n фрагментів, в результаті чого сам набір також діліться на n частин Q_i, $i = 1, 2, \ldots n$. Після синтезу в процесі функціонування кожний каскад КС$_i$ в складі МР виконує розпізнавання відповідних фрагментів патернів в потоці вхідних даних, що подаються одночасно на входи всіх КС, але лише у випадку виявлення збігу для відповідного фрагменту з попереднього каскаду. Сигнали про частковий збіг передаються з попереднього на наступний каскад. В разі виявлення збігу вхідної послідовності з якимось фрагментом цілком, в останньому каскаді КС$_n$, активується відповідний сигнал розпізнавання.

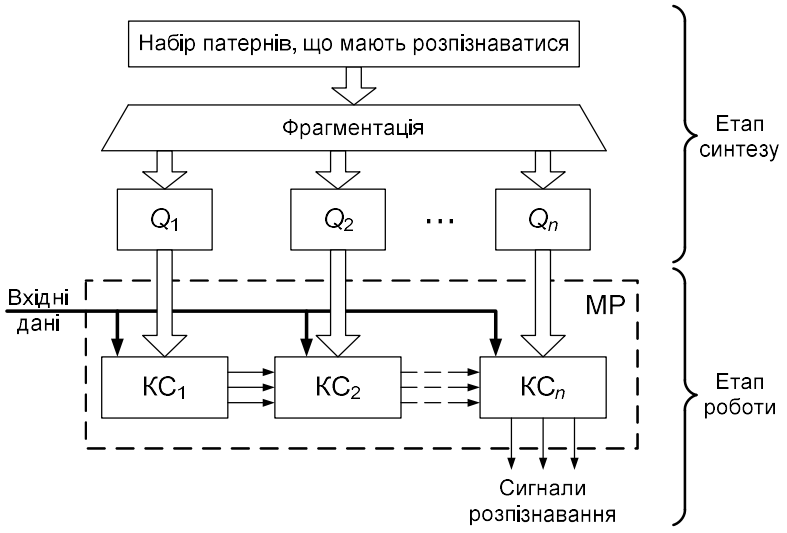


Рис. 9. Схематичне подання структури методу послідовного каскадування

Виходячи з припущення, що збіг вхідних даних з довшим фрагментом патерну є менш вірогідною подією, ніж з коротшим, можна очікувати, що потреба у використанні кожного наступного каскаду буде виникати рідше порівняно з попереднім, внаслідок чого вимоги до швидкодії (та, як наслідок – споживаних ресурсів) каскадів знижатимуться за зростанням їх номеру аж до можливості реалізації останніх каскадів програмними засобами. Водночас за рахунок розпізнавання більш коротких підрядків для перших каскадів може бути досягнута вища швидкодія. Отже побудова МР за каскадною схемою призводить одночасно як до підвищення швидкодії РАСЗТЗІ, так і до скорочення потрібних ресурсів.

До другого і наступних каскадів висуваються менш жорсткі вимоги не тільки за швидкодією, але й за функціональністю. Тому при їх побудові можуть

використовуватися не тільки повноцінні підходи, але й базові технології (наприклад, окремі ЦК без створення з них АП чи геш-функції без побудови ФБ), що також є перевагою методу МПсКс.

Для знаходження ефективного розбиття патернів на фрагменти для даного методу також доцільно використовувати процедуру оптимізації.

Особливістю методу МПсКс є необхідність крім фіксованих властивостей питомій підгрупи патернів враховувати статистичні характеристики вхідного інформаційного потоку, зокрема, у випадку застосування МСВВ – імовірнісні характеристики мережевого трафіку. Це обумовлено тим фактом, що частота запуску наступних каскадів і, як наслідок, вимоги до їх швидкодії, залежать від властивостей вхідних даних. Слідством цієї особливості є головний недолік методу МПсКс – непередбачуваність пропускної здатності та нездатність в окремих випадках протидіяти атакам, спрямованим на засоби захисту інформації. Цей факт треба враховувати при застосуванні даного методу для побудові МР РАСЗТЗІ.

Принцип дії ***Методу вертикального об'єднання*** (МВрОб), розвинутого автором, полягає в сполучанні в одному модулі кількох підходів та окремих технологій таким чином, що жоден з них не може бути відокремленим від інших. Зазвичай при такому об'єднанні один з підходів, під який будується схема в цілому, виконує основну функцію, а інші вирішують допоміжні задачі.

На рис. 10 наведено схематичне зображення структури МВрОб. Як можна бачити, при використанні даного методу набір патернів, що розпізнається, ніяким чином не діліться на підгрупи. Тобто, побудований за методом МВрОб модуль діє як єдиний функціональний блок.

Таблиця сумісності для МВрОб. З метою систематизації наявного досвіду та спрощення процесу розробки компонентів РАСЗТЗІ методом МВрОб, автором запропонований його розвиток, суть якого полягає в створенні та наповненні за певними правилами інформаційної структури, призначеної для зберігання та використання відомостей, отриманих емпіричним шляхом. Назвемо цю структуру багатовимірною таблицею сумісності вертикального об'єднання (БТСВО).

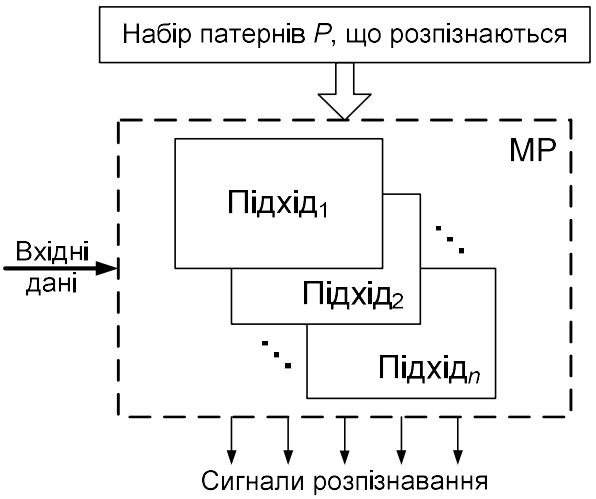


Рис. 10. Схематичне подання структури методу вертикального об'єднання

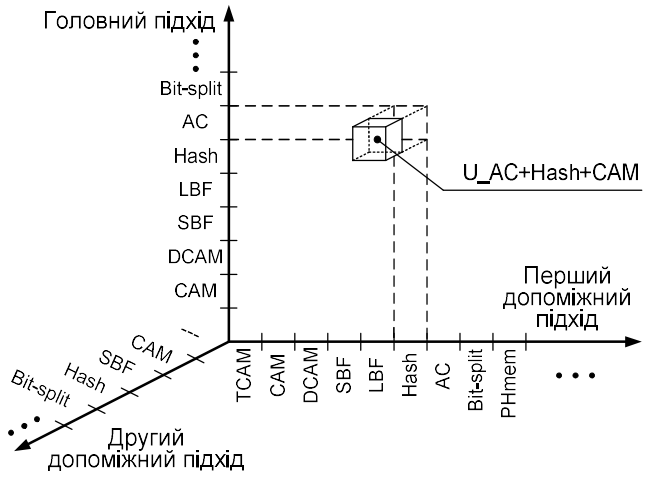


Рис. 11. Багатовимірна таблиця сумісності вертикального об'єднання

Технічно БТСВО організована у вигляді таблиці (рис. 11), вздовж кожного

виміру якої розташовані всі відомі підходи, їх модифікації та технології, які теоретично здатні бути складовими частинами для сполучення в єдину схему розпізнавання за принципом МВрОб. Кожна комірка, що розташована на перехресті всіх вимірів, містить наступну інформацію стосовно об'єднання відповідної комбінації аргументів: чи можуть обрані техніки/технології функціонувати спільно, тобто, чи існує принципова можливість об'єднати відповідні складові в один модуль розпізнавання методом МВрОб; якщо так, в якому порядку відбудовуватиметься їх взаємодія, тобто, який напрям буде основним, а які допоміжними; які характеристики основного рішення та на який коефіцієнт покращуватимуться допоміжним техніками (мультиплікаторами) в результаті об'єднання, якщо такі коефіцієнти існують для даної комбінації аргументів.

Після заповнення БТСВО наявною інформацією значна частина її комірок залишиться порожніми, тобто в них не буде жодних відомих розробникам даних щодо можливості об'єднання складових, що в ній перетинаються. Це прискорює процедуру оптимізації у випадку використання методу повного перебору.

Комбінування методів комбінування. Аналізуючи досліджені вище методи комбінування, можна виявити суттєву різницю між їх природою та взаємовідношеннями між технічними рішеннями, що комбінуються. У кожного з цих методів є переваги та недоліки, що природним чином приводить до ідеї про їх комбінування один з одним, тобто про комбінування методів комбінування (КМК).

На рис. 12 наведено приклад одночасного комбінування всіх трьох методів.

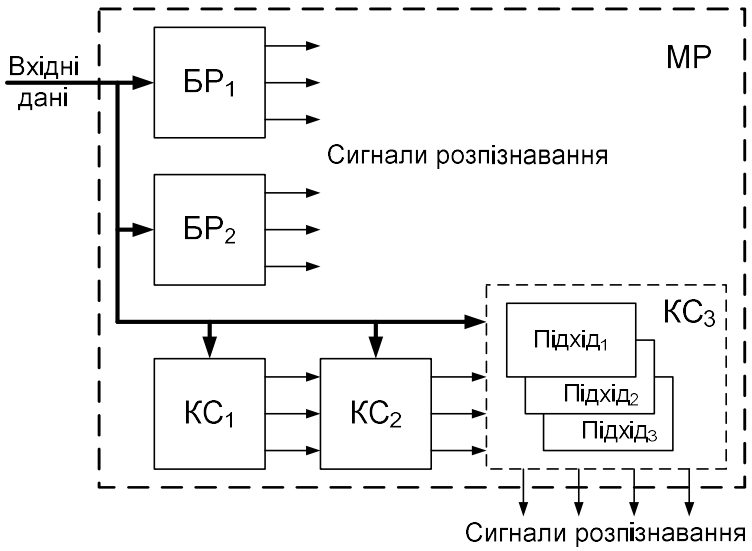


Рис. 12. Приклад комбінування методів комбінування

Як бачимо, загальна структура отриманої комбінації в цілому нагадує метод паралельного комбінування, але один з блоків розпізнавання замінено на структуру методу послідовного каскадування, один с каскадів якого в свою чергу є реалізацією метода вертикального об'єднання. Практична реалізації принципу КМК не викликає труднощів, тому що відповідні алгоритми просто вкладаються один в одного згідно заданої ієрархії.

Теоретично, згідно принципу КМК може створюватися ієрархія методів комбінування довільної глибини. Але в реальних застосуваннях глибина більша за три рівні, фактично, не використовується.

У шостому розділі автором запропоновано алгоритми реалізації методів комбінування, розроблено програмні засоби та проведено експерименти.

Згідно визначенню методу МПрКм, сформульованому в попередньому розділі, алгоритм побудови МР РАСЗТЗІ за цим методом має містити три вкладених цикли:

1. Цикл вибору кількості БР – варіюється кількість n блоків розпізнавання $БР_i$ в діапазоні від 2 до n_{max}, де n_{max} – максимальне значення n, задане користувачем.

2. Цикл вибору БР – з бібліотеки компонентів обираються схеми БР$_i$, побудовані за різними підходами, модифікаціями підходів та техніками.

3. Цикл розбиття патернів – варіюються різні комбінації розбиття словнику патернів P на n наборів.

Перший та другий цикли нескладно реалізувати методом повного перебору внаслідок невеликого числа змінних. З третім циклом ситуація виявляється складнішою. Кількість варіантів розбиття множини патернів P, що налічує σ патернів $p_1, p_2, \dots p_k, \dots p_\sigma$ між n блоками розпізнавання БР1, БР2,… БР$_i$, … БР$_n$ дорівнює $W_\sigma = n^\sigma$. Для баз даних сигнатур сучасних РАСЗТЗІ, які налічують до сотень тисяч патернів, кількість варіантів поділу W_σ навіть між двома БР$_i$, є комбінаторно великим числом, що унеможливлює застосування повного перебору. Дещо знизити цю кількість дозволяє оперування не окремими патернами, а пакетами патернів однакової довжини. (Якщо знехтувати ефектом самоподоби, розпізнавання патернів однакової довжини призводить до однакових значень функції оцінки). Тоді кількість варіантів розбиття множини патернів P, що налічує ξ пакетів по δ_j патернів однакової довжини, між блоками розпізнавання в кількості n одиниць знизиться до $W_\xi = n^\xi$. Але це зменшує ступінь числа блоків розпізнавання n лише на один – два десяткові порядки. Тобто число W_ξ все рівно залишається завеликим для практичної реалізації. Використання інших методів оптимізації, наприклад, покоординатного спуску, не є можливим в зв'язку з відсутністю будь-якої метрики в просторі варіантів, який є комбінацією змінних.

З метою усунення комбінаторної складності обчислень шляхом надання множині комбінацій метричних здібностей автором запропоновано ***Метод прискорення процедури оптимізації паралельного комбінування***, суть якого полягає у впорядкуванні патернів в наборі за певним параметром. Тоді замість повного перебору можна варіювати лише цей параметр вздовж діапазону можливих значень, внаслідок чого обсяг обчислень істотно зменшується. Хоча такий підхід не гарантує знаходження глобального оптимуму, проте за рахунок скорочення об'єму обчислень з комбінаторно великого до лінійного закону він дозволяє реалізувати процедуру оптимізації за прийнятний час.

Реалізація методу призводить до наступного алгоритму. Для $n = 2$ на першому кроці блокові БР1 призначуються всі патерни набору P, тоді як блокові БР2 – жодного, тобто пусту множину патернів. На наступному кроці блокові БР1 призначуються всі патерни набору крім першого патерну за обраним параметром сортування, який надається для розпізнавання блокові БР2. На наступному кроці вже два патерни – перший та другий за обраним параметром сортування переходять від БР1 до блоку БР2. Процедура повторюється до ситуації, коли в підмножині, яку розпізнає блок БР1 не залишається жодного патерну, а блок БР2 отримує для розпізнавання весь набір патернів P. На кожному кроці обчислюються значення ФО для кожного БР та для комбінованої структури в цілому. У випадках, коли кількість блоків n більша за 2, алгоритм можна скласти рекурсивно, розглядаючи один з блоків як такий, що складається з двох інших блоків.

В даному дослідженні автором запропоновано три типи впорядкування: за

довжиною патернів (Pattern_L), за кількістю патернів у пакеті (Delta_j) та за "площею" пакета (Pat_x_Del), тобто за добутком згаданих величин.

На рис. 13 наведено схематичне подання використання запропонованих в даному дослідженні методів та алгоритмів.

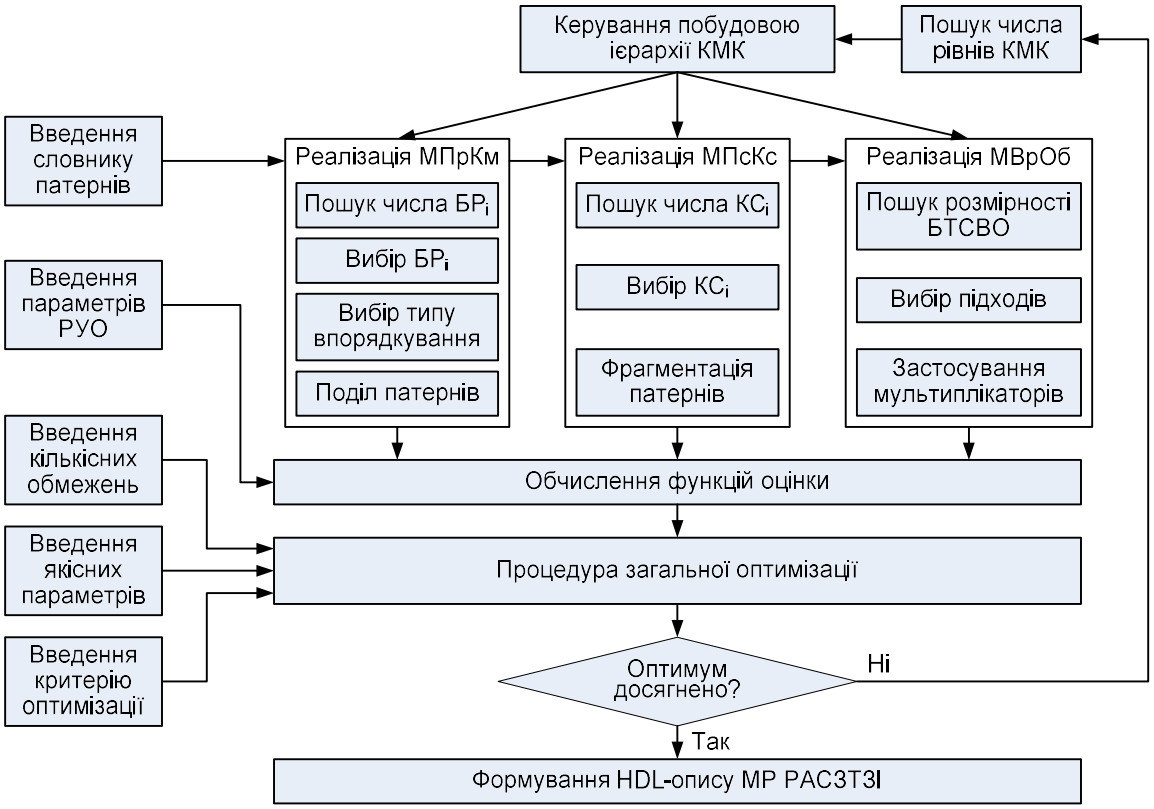
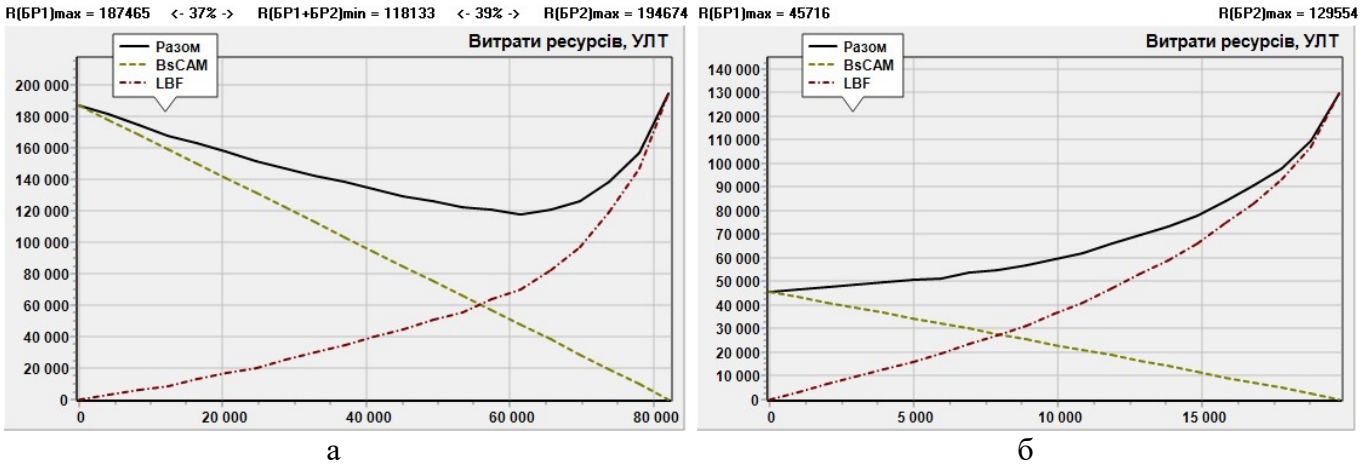


Рис. 13. Загальна структура використання методів комбінування

На рис. 14 а) наведено результати розрахунків для прикладу мінімізації ресурсних витрат методом МПрКм за наступних умов: $n = 2$, реалізація $БР_1$ – за схемою BsCAM, $БР_2$ – за схемою LBF, впорядкування – Pat_x_Del (за спаданням). Набір патернів P був використаний з вільно розповсюджуваної бази даних сигнатур "Community Ruleset" відкритої МСВВ Snort версії 3.0, що містить $\sigma = 4208$ патернів довжиною від $m_{\min} = 1$ до $m_{\max} = 364$ загальною кількістю $\Omega = 82081$ символ.



а б

Рис. 14. Комбінування методом МПрКм схем розпізнавання BsCAM і LBF: а – вдале; б – невдале

По осі абсцис тут відображається кількість символів у патернах, які передані на розпізнавання від блоку БР1 до блоку БР2. По осі ординат – витрати в умовних ЛТ на створення кожного з БР та разом на МР, що з них складається. Як можна бачити, крива витрат на МР має виражений мінімум, при якому витрати загалом на весь МР на 37% менші витрат на БР1, та на 39% – на БР2.

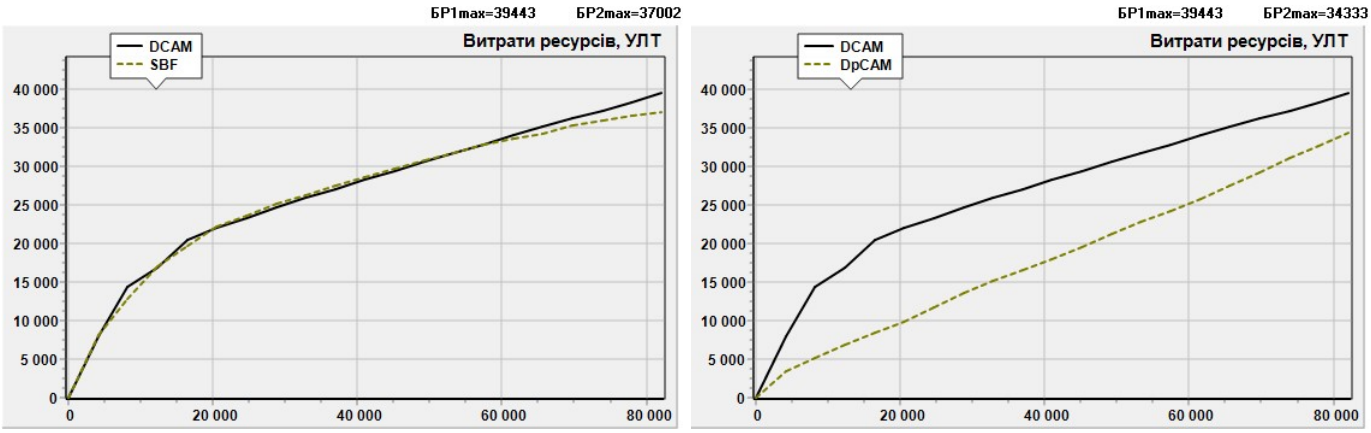
На практиці вдалий результат паралельного комбінування, подібний до наведеного на рис. 14 а), виходить рідко. Набагато частіше один з блоків, що комбінуються, вздовж всього виміру варіювання демонструє кращі значення окремо, ніж сумісно з іншим. Подібний приклад наведено на рис. 14 б). Тут задіяні ті ж самі схеми розпізнавання, але використаний інший набір патернів – "Web_server.rules" (атаки на WEB-сервери) зі складу бази даних сигнатур іншої відкритої МСВВ Suricata версії 5.0. Набір містить 999 патернів довжиною від 1 до 233 загальною кількістю 19752 символів.

За таких умов витрати на створення БР1 у граничному випадку, тобто коли він розпізнає весь набір патернів, в декілька разів менші, ніж на створення БР2 в разі розпізнавання ним того ж самого набору патернів.

Крім ситуації, описаної вище, можливість використання методу МПрКм також обмежує необхідність узгодження швидкісних характеристик $БР_i$.

Але, як свідчать експерименти, навіть якщо блоки розпізнавання мають однакову швидкодію та близькі ресурсні витрати у граничних випадках, їх паралельне комбінування все рівно не матиме позитивного ефекту, якщо функціональні залежності їх характеристик від поділу патернів подібні.

На рис. 15 а) наведено порівняння кривих ресурсних витрат для модифікації підходу на базі цифрових компараторів DCAM та одного з варіантів реалізації схеми фільтра Блума SBF (для бази даних сигнатур "Community Ruleset" при впорядкуванні за типом Pat_x_Del). Як можна бачити, характер поведінки обох блоків розпізнавання вздовж всього виміру варіювання відрізняється не суттєво. Тому паралельне комбінування цих двох технічних рішень гарантовано не призведе до появи мінімуму сумарної кривої, як на рис. 14 а).



а б

Рис. 15. Порівняння схем, побудованих за різними та спорідненими підходами: а – DCAM і SBF (впорядкування Pat_x_Del); б – DCAM і DpCAM (впорядкування Pat_x_Del)

Такий результат експерименту розбігається з теоретичним припущенням щодо відмінності властивостей схем розпізнавання, побудованих за різними підходами.

Більш того, виявляється, що, з іншого боку, модифікації одного підходу можуть істотно розрізнятися між собою. На рис. 15 б) зведено разом характеристики двох модифікацій DCAM та DpCAM одного й того же підходу на базі ЦК. Як можна бачити, попри близькість внутрішніх структур цих схем (вони є відповідно декодованою та частково декодованою модифікаціями базової схеми BsCAM на ЦК), характер поведінки в них суттєво різний.

Факти розбіжності теоретичних міркувань з результатами розрахунків свідчать про важливість експериментальних досліджень для даної роботи.

З метою скорочення числа варіантів комбінування для прискорення методу МПрКм доцільно заздалегідь порівняти між собою поведінку наявних БР$_i$ (рис. 16) за різними типами впорядкування патернів та виключити з розглядання під час оптимізації сумісне використання таких блоків, що мають подібну поведінку.

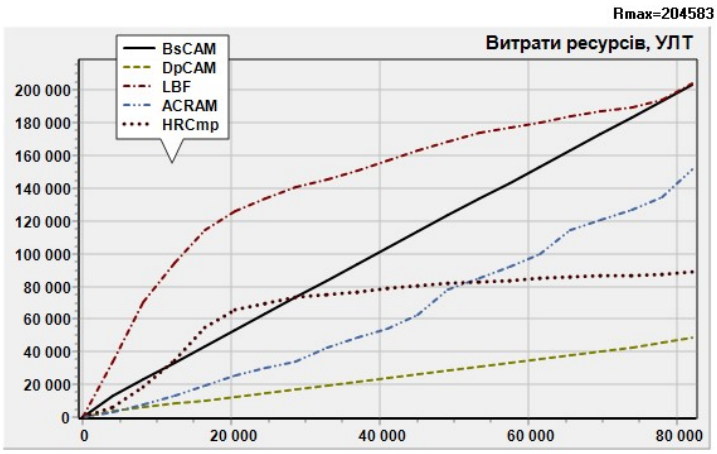


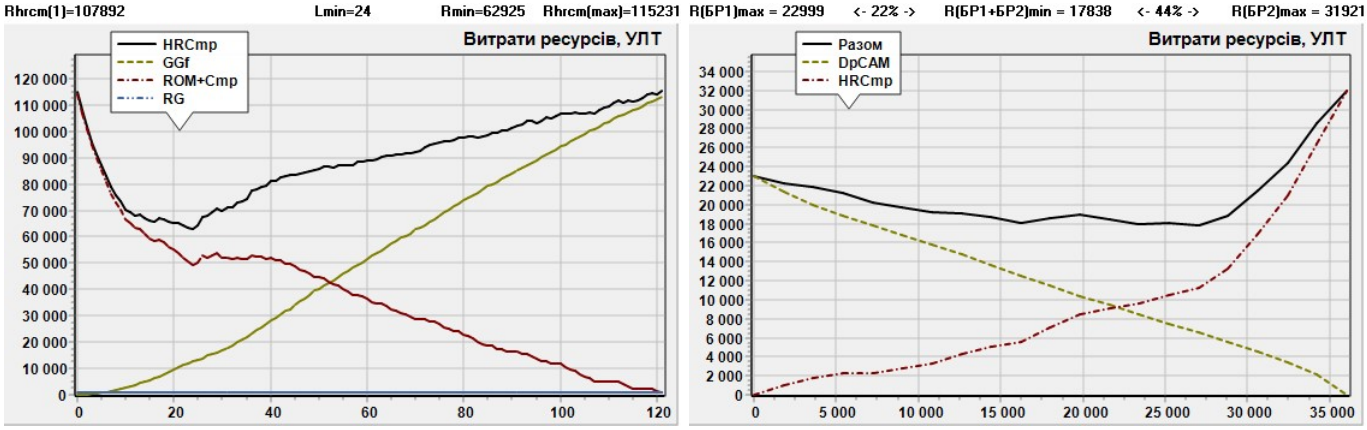
Рис. 16. Попереднє порівняння декількох схем розпізнавання (впорядкування типу Pattern_L за спаданням)

Завершуючи розгляд експериментального дослідження методу МПрКм, зауважимо, що навіть у випадку, коли не вдається отримати виграш від паралельного комбінування декількох різнорідних схем, тому що один з БР попереджає всі інші, головну мету все одно досягнуто – виявлено найбільш ефективну за наданих умов схему розпізнавання для МР РАСЗТЗІ. А отриманий результат можна трактувати як вироджений випадок МПрКм для $n = 1$, який швидко здобуто завдяки застосуванню методів прискорення.

Алгоритм реалізації методу МПсКс простіше, ніж для МПрКм, тому що не потребує перебору комбінаторно великої кількості варіантів. Для знаходження оптимуму для даного методу потрібно знайти в загальному випадку всього $n - 1$ питому змінну – номера символів, по яких відбуватиметься фрагментація патернів. У найчастішому випадку, коли $n = 2$, метод вироджується в пошук тільки однієї величини – довжини префіксів, що розпізнаватимуться першим каскадом.

Експериментальні дослідження підтвердили, що для тих схем розпізнавання, що комбінуються методом МПсКс, оптимум впевнено знаходиться. На рис. 17 а) наведено приклад залежності від довжини префіксу L ресурсних витрат на побудову двокаскадної схеми комбінування HRCmp, яка в першому каскаді використовує хешування (без створювання повноцінної схеми ФБ), а в другому – постійний запам'ятовуючий пристрій та компаратори (без створення повної схеми АП). Тут по осі абсцис відкладено довжину префікса патернів L у символах, що розпізнаватимуться першим каскадом. Літерами GGf позначена зростаюча крива витрат на генератор геш-функції, ROM+Cmp – сумарні витрати на запам'ятовуючий пристрій та компаратори, RG – на конвеєр вхідної послідовності символів. Впорядкування тут не потрібне, тому що природну метрику створює довжина префікса патернів L (довжина суфіксів патернів при цьому дорівнює $(m_j - L)$,

де m_j – довжина j-го патерну). В даному випадку оптимум досягається при значенні $L = 24$ символи.



а б

Рис. 17. Використання методу МПсКс: а – склад ресурсних витрат схеми HRCmp;
б – паралельне комбінування DpCAM та HRCmp

Зауважимо, що в наведеному прикладі вдалося ефективно задіяти переваги технології гешування завдяки тому, що в першому каскаді розпізнається велика кількість патернів (точніше префіксів) однакової довжини, а, як відомо, основний недолік рішень на базі геш-функцій пов'язаний з різною довжиною патернів.

Оскільки при застосуванні методу МПсКс в більшості розробок використовуються не повноцінні схеми розпізнавання, а лише базові технології, коректно оцінити його ефективність складно за відсутністю рівноцінних аналогів. Але порівняння з повноцінними схемами розпізнавання, побудованими за тими ж технологіями, що використані при комбінуванні за методом МПсКс демонструє перевагу в зниженні ресурсних витрат на 40–45% (див., наприклад, рис. 17 а) при збереженні швидкодії, проте – лише на відносно коротких патернах.

Тому виникає природний інтерес щодо об'єднання в одному пристрої методів МПсКс та МПрКм за принципом КМК, розглянутому в попередньому розділі.

На рис. 17 б) наведено результати розрахунків ресурсних витрат для модулю розпізнавання, в якому паралельно скомбіновано схему DpCAM та схему HRCmp, яка в свою чергу є результатом послідовного комбінування. Як можна бачити, за наданих умов такий модуль споживає на 22% менше ресурсів порівняно зі схемою DpCAM, та на 44% порівняно зі схемою HRCmp. Даний приклад цікавий тим, що подібну схему було реалізовано та досліджено у низці відомих публікацій, автори котрих не використовували формалізований апарат методів комбінування, але інтуїтивно застосовували прийоми, що призвели до схожого результату. Вони відсортували всі патерни за довжиною та "застосували схему HRCmp для розпізнавання патернів довжиною до 50 символів та DpCAM для довших патернів", що призвело до покращення ПЕ на 10–25% [I. Sourdis, D. N. Pnevmatikatos, and S. Vassiliadis, "Scalable multigigabit pattern matching for packet inspection," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, Article vol. 16, no. 2, pp. 156-166, Feb 2008]. Наведений на рис. 17 б) приклад розраховано за тих самих умов, що були використані в згаданих публікаціях (словник патернів з бази даних сигнатур Snort 2007 року версії 2.3.2, що містить $\sigma = 2188$ патернів довжиною від

$m_{\min} = 1$ до $m_{\max} = 122$, загальною кількістю $\Omega = 33618$ символів). Але за рахунок використання формалізованих методів підвищення ефективності з використанням оптимізації та прискореного обчислення технічних характеристик в даному дослідженні вдалося з'ясувати, що в даному випадку можна отримати кращі результати (на 12–19%), якщо поділити патерні на значенні довжини в 45 символів.

У сьомому розділі розглянуто експериментальну систему централізованого синтезу компонентів РАСЗТЗІ та програмування ПЛІС реконфігуровних обчислювачів з використанням засобів високопродуктивного обчислень.

Процес розробки обчислювальної структури реконфігуровних пристроїв є складною та ресурсномісткою задачею, що вимагає від розробника володіння на високому рівні засобами САПР, знання специфіки проектування цифрових пристроїв, а від комп'ютерної техніки – високої продуктивності. Користувачі РАСЗТЗІ (системні адміністратори та персонал, відповідальний за інформаційну безпеку) не мають умов для самостійного вирішення цієї задачі. автором було запропоновано організувати процес синтезу реконфігуровних засобів таким чином, щоб складні та ресурсномісткі процедури виконувалися не локально на кожній окремій системі, а централізовано, з використанням високопродуктивних комп'ютерних технологій, таких як грід-мережі та хмарні обчислення.

Технічно принцип централізованого синтезу був реалізований у вигляді макетного зразка веб-сервісу STRAGS (Security Tasks Reconfigurable Accelerators Grid-Service) на базі кластера Інституту проблем моделювання в енергетиці ім. Г.Є. Пухова НАН України за безпосередньою участю автора. Інтерфейс сервісу дозволяє користувачеві завантажити завдання зі своїми даними (переліком сигнатур, типом використаного РУО та додатковими вимогами щодо процесу розпізнавання). По завершенню виконання завдання клієнт сервісу отримує результати обчислень у вигляді файлів конфігурації для ПЛІС та лог-файли зі службовою інформацією.

Як було з'ясовано у другому розділі, під час проведення ПОО технологічний ланцюжок синтезу МР РАСЗТЗІ зводиться до двох етапів – формування машинних описів змінних компонентів та генерації конфігурації для ПЛІС. На першому етапі, більш наукоємному, по змісту бази даних сигнатур РАСЗТЗІ генерується опис схеми класифікатора заголовків та формується словник патернів. Процедура реалізації модулю розпізнавання виконується із застосуванням методів та засобів, створених автором в даному дослідженні. До згенерованих блоків змінної складової – класифікатора та МР, додаються схемні заготовки постійної складової компонентів. Описи всіх підсхем створюються мовою опису апаратури – HDL (Hardware Definition Language). На другому етапі, більш ресурсоємному обчислювально, HDL-опис доповнюється файлами обмежень (ucf-файлами) та іншими складовими проекту створення цифрового пристрою, після чого починає роботу процедура генерації конфігурацій із застосуванням пакета САПР виробника ПЛІС, використовуючи обчислювальні ресурси Українського національного гріду та хмарного сервісу Amazon AWS.

В процесі функціонування сервісу STRAGS його грідівська складова ініціює роботу на віддалених вузлах грід-середовища декількох агентів – віртуальних машин з попередньо встановленим і налаштованим потрібним програмним забезпеченням, до якого входять, по-перше, програмно реалізовані алгоритми

запропонованих автором методів комбінування, по-друге, інструментальні пакети створення файлів конфігурацій за їх HDL-описами. По надходженню запитів від клієнтів сервіс розподіляє завдання між активними агентами, підтримуючи їх число достатнім для забезпечення готовності на необхідному рівні. Отримавши задачу у вигляді грід-завдання, агент запускає потрібні обчислювальні процеси, після чого повертає результати роботи сервісу. В якості хмарної складової сервісу STRAGS задіяний веб-сервіс Amazon Web Services. Віртуальні машини – хмарні агенти запускаються на рівні IaaS (Infrastructure as a Service) цього сервісу.

У висновках зроблено підсумок результатів дисертаційної роботи.

ВИСНОВКИ

В дисертаційній роботі, на основі проведених досліджень, вирішена важлива науково-технічна проблема, що полягає в розробці та розвитку методів побудови комбінованих обчислювальних структур для підвищення ефективності реконфігуровних сигнатурних технічних засобів захисту інформації.

Під час дослідження були отримані наступні наукові та практичні результати:

1. Проведено аналіз проблем технічного захисту інформації, які вирішуються сигнатурними засобами, що дозволило сформулювати задачу множинного розпізнавання рядків та ввести поняття процедури оперативної оновлення.
2. Проведено дослідження технічних можливості сучасних ПЛІС та пристроїв на їх основі в якості апаратної бази для вирішення задач технічного захисту інформації, що дозволило порівняти її з іншими апаратними платформами.
3. Проведено аналіз принципів функціонування та побудови РАСЗТЗІ як технічних систем. Сформульовано та класифіковано показники їх ефективності, що дало можливість оцінювати та порівнювати ефективність як окремих компонентів створюваних засобів захисту інформації, так и системи в цілому.
4. В результаті вивчення та дослідження наявного світового досвіду побудови реконфігуровних сигнатурних засобів захисту, виокремлено три найбільш ефективні підходи до побудови модуля розпізнавання РАСЗТЗІ; шляхом дослідження цих підходів та їх модифікацій на структурному рівні формалізовано властивості кожного з них, враховуючи специфіку реалізації реконфігуровними засобами, що дозволило сформулювати у вигляді наукових методів ідею комбінування (сумісного використання) різних підходів до побудови компонентів РАСЗТЗІ з метою використання переваг кожного з підходів.
5. Розроблено метод прискореного обчислення технічних характеристик на основі функцій оцінки, який на відміну від відомих методів проектування РАСЗТЗІ не потребує виконання витратних за часом процедур синтезу цифрових схем, що дозволило здійснювати швидку оцінку за заданими показниками ефективності окремих компонентів і РАСЗТЗІ в цілому.
6. Розроблено метод прискорення процедури оптимізації паралельного комбінування, суть якого полягає у впорядкуванні патернів за певним параметром, що дозволило за рахунок послідовної зміни параметру впорядкування пришвидшити поділ патернів між блоками, що комбінуються, від повного перебору комбінаторно великої кількості варіантів до витрат часу за

лінійним законом. Метод не гарантує знаходження глобального оптимуму, проте зменшує обчислювальну складність процедури оптимізації до прийнятних для практичного використання значень від десятків секунд до десятків хвилин.

7. На основі розроблених методів прискорення обчислень, виявлено функціональні залежності кількісних характеристик найбільш ефективних схем розпізнавання від параметрів реконфігуровного обчислювача та властивостей заданого набору патернів, що дозволило здійснити експериментальну перевірку методів комбінування.
8. Сформульовано та вдосконалено метод паралельного комбінування, який за рахунок паралельного з'єднання різних за принципами побудови блоків розпізнавання та оптимізації поділу між ними набору патернів, використовуючи розроблені методи прискорення, дозволяє підвищити показники ефективності модулю розпізнавання РАСЗТЗІ в окремих випадках на 35-40% порівняно з блоками розпізнавання, побудованими з використанням одного з підходів.
9. Сформульовано та вдосконалено метод послідовного каскадування, який за рахунок послідовно з'єднання різних за принципами побудови блоків розпізнавання та оптимізації поділу між ними питомих патернів по довжині, використовуючи розроблені методи прискорення, дозволяє підвищити показники ефективності модулю розпізнавання РАСЗТЗІ в окремих випадках на 40-45% порівняно з блоками розпізнавання, побудованими з використанням одного з підходів.
10. Отримав подальший розвиток метод вертикального об’єднання, який за рахунок тісного сполучання в одному блоці кількох підходів або технічних рішень дозволяє підвищити показники ефективності модулю розпізнавання РАСЗТЗІ, а саме, запропоновано модифікацію метода, яка за рахунок використання багатовимірної таблиці сумісності дозволяє формалізувати наявний досвід розробників РАСЗТЗІ та спростити процедуру пошуку найбільш ефективної за наданих умов комбінації задіяних даним методом підходів або технічних рішень.
11. Сформульовано та вдосконалено принцип комбінування методів комбінування, який за рахунок ієрархічного використання сформульованих методів комбінування дозволяє підвищити показники ефективності модулю розпізнавання РАСЗТЗІ в окремих випадках на 20-45% порівняно з використанням окремих методів комбінування.
12. Розроблено алгоритми реалізації вдосконалених та розвинутих методів комбінування, що дозволило створити програмні засоби їх перевірки та аналізу.
13. Розроблено програмні засоби для перевірки та аналізу вдосконалених та розвинутих методів, що дозволило шляхом проведення обчислювальних експериментів порівняти їх з відомими розробками, в яких застосовуються методи комбінування, та з'ясувати, що запропоновані та вдосконалені методи за рахунок прискорення обчислень та використання процедур оптимізації дозволяють перевершити відомі розробки в окремих випадках на 12-19%.
14. Досліджено та розвито принцип централізованого створення обчислювальних структур реконфігуровних пристроїв, що дозволило розробити на його основі веб-сервіс, що реалізує запропоновані методи та засоби, використовуючи розподілені та хмарні середовища.

СПИСОК ОПУБЛІКОВАНИХ ПРАЦЬ ЗА ТЕМОЮ ДИСЕРТАЦІЇ

Наукові праці, в яких опубліковані основні наукові результати дисертації:

1. Гильгурт С.Я. Применение реконфигурируемых вычислений для решения задач распознавания / С.Я. Гильгурт // Зб. наук. пр. ІПМЕ ім. Г.Є. Пухова НАН України. – Київ, 2002. – Вип. 16. – С.41-47.

2. Гильгурт С.Я. Анализ применений программируемых ИС для решения задач вычислительной техники / С.Я. Гильгурт // Зб. наук. пр. ІПМЕ ім. Г.Є. Пухова НАН України. – Київ, 2003. – Вип. 19. – С.79-85.

3. Гильгурт С.Я. Особенности применения реконфигурируемых вычислителей для аппаратной защиты информационных систем / С.Я. Гильгурт // Зб. наук. пр. ІПМЕ ім. Г.Є. Пухова НАН України. – Київ, 2007. – Вип. 38. – С.36-41.

4. Гильгурт С.Я. О применении реконфигурируемых вычислителей для решения задач защиты информации / С.Я. Гильгурт, А.К. Гиранова // Зб. наук. праць ІПМЕ ім. Г.Є. Пухова НАН України. – Київ, 2008. – Вип. 46. – С.93-99.

5. Гильгурт С.Я. Обзор современных реконфигурируемых унифицированных вычислителей / С.Я. Гильгурт // Моделювання та інформаційні технології. Зб. наук. пр. ІПМЕ ім. Г.Є. Пухова НАН України. – Київ, 2008. – Вип. 49. – С.17-24.

6. Гильгурт С.Я. Анализ применения унифицированных вычислителей в интеллектуальных системах / С.Я. Гильгурт // Искусственный интеллект. – 2009. – № 1. – С.144-148.

7. Гильгурт С.Я. О применении реконфигурируемых вычислителей для решения задач защиты информации / С.Я. Гильгурт, А.К. Гиранова // Зб. наук. праць ІПМЕ ім. Г.Є. Пухова НАН України. – Київ, 2009. – Вип. 51. – С.65-72.

8. Коростиль Ю.М. Анализ угроз и опасностей в компьютерных системах на предмет защиты цифровыми реконфигурируемыми устройствами / Ю.М. Коростиль, А.Н. Давиденко, С.Я. Гильгурт, М.М. Панченко // Моделювання та інформаційні технології. Зб. наук. пр. ІПМЕ ім. Г.Є. Пухова НАН України. – Київ, 2010. – Вип. 56. – С.10-17.

9. Гильгурт С.Я. Обзор возможностей реконфигурируемых устройств для применения в компьютерной безопасности / С.Я. Гильгурт // Зб. наук. пр. ІПМЕ ім. Г.Є. Пухова НАН України. – Київ, 2010. – Вип. 55. – С.117-124.

10. Коростиль Ю.М. Принципы построения сетевых систем обнаружения вторжений на базе ПЛИС / Ю.М. Коростиль, С.Я. Гильгурт // Моделювання та інформаційні технології. Зб. наук. пр. ІПМЕ ім. Г.Є. Пухова НАН України. – Київ, 2010. – Вип. 57. – С.87-94.

11. Давиденко А.Н. Алгоритмы распознавания строк в системах обнаружения вторжений на ПЛИС / А.Н. Давиденко, С.Я. Гильгурт // Моделювання та інформаційні технології. Зб. наук. пр. ІПМЕ ім. Г.Є. Пухова НАН України. – Київ, 2010. – Вип. 58. – С.103-109.

12. Гильгурт С.Я. Программно-аппаратная защита данных в распределенных интеллектуальных системах / С.Я. Гильгурт, А.К. Гиранова // Искусственный интеллект. – 2010. – № 3. – С.706-711.

13. Гильгурт С.Я. Анализ типовых режимов обмена данными с реконфигурируемыми вычислителями / С.Я. Гильгурт // Зб. наук. пр. ІПМЕ

ім. Г.Є. Пухова НАН України. – Київ, 2011. – Вип. 59. – С.113-121.

14. Гильгурт С.Я. Аппаратное распознавание строк в интеллектуальных системах защиты информации / С.Я. Гильгурт // Искусственный интеллект. – 2012. – № 1. – С.259-266.

15. Гильгурт С.Я. Сравнительный анализ межсетевых экранов, систем обнаружения и предотвращения вторжений, контроля целостности / С.Я. Гильгурт // Зб. наук. пр. ІПМЕ ім. Г.Є. Пухова НАН України. – Київ, 2011. – Вип. 60. – С.116-120.

16. Давиденко А.Н. Аппаратное ускорение алгоритмов сигнатурного обнаружения вторжений в открытой системе информационной безопасности Snort / А.Н. Давиденко, С.Я. Гильгурт, В.И. Сабат // Моделювання та інформаційні технології. Зб. наук. пр. ІПМЕ ім. Г.Є. Пухова НАН України. – Київ, 2012. – Вип. 65. – С.94-103.

17. Коростиль Ю.М. Анализ базы данных системы информационной безопасности Snort и вопросы быстродействия / Ю.М. Коростиль, С.Я. Гильгурт, О.М. Назаренко // Моделювання та інформаційні технології. Зб. наук. пр. ІПМЕ ім. Г.Є. Пухова НАН України. – Київ, 2012. – Вип. 66. – С.77-84.

18. Коростиль Ю.М. Перспективы развития реконфигурируемых вычислителей для выполнения ресурсоемких расчетов / Ю.М. Коростиль, С.Я. Гильгурт // Моделювання та інформаційні технології. Зб. наук. пр. ІПМЕ ім. Г.Є. Пухова НАН України. – Київ, 2013. – Вип. 68. – С.84-91.

19. Гильгурт С.Я. Реконфигурируемые вычислители. Аналитический обзор / С.Я. Гильгурт // Электронное моделирование. – 2013. – Т. 35, № 4. – С.49-72.

20. Гильгурт С.Я. Противодействие атакам алгоритмической сложности на системы обнаружения вторжений / С.Я. Гильгурт, Б.В. Дурняк, Ю.М. Коростиль // Моделювання та інформаційні технології. Зб. наук. пр. ІПМЕ ім. Г.Є. Пухова НАН України. – Київ, 2014. – Вип. 71. – С.3-12.

21. Hilhurt S. Ya. Application of FPGA-based Reconfigurable Accelerators for Network Security Tasks / S. Hilhurt // Collection of scientific works. Simulation and informational technologies. – PIMEE NAS of Ukraine. – Kyiv, 2014. – Vol. 73. – P. 17–26.

22. Гильгурт С.Я. Организация вычислительного процесса синтеза файлов конфигураций для аппаратных ускорителей при решении задач информационной безопасности / С.Я. Гильгурт // Моделювання та інформаційні технології. Зб. наук. пр. ІПМЕ ім. Г.Є. Пухова НАН України. – Київ, 2015. – Вип. 74. – С.29-33.

23. Євдокимов В.Ф. Створення на базі грід-сайту ІПМЕ ім. Г.Є. Пухова НАНУ системи централізованого синтезу апаратних прискорювачів для вирішення задач інформаційної безпеки в енергетичній галузі / В.Ф. Євдокимов, А.М. Давиденко, С.Я. Гільгурт // Моделювання та інформаційні технології. Зб. наук. пр. ІПМЕ ім. Г.Є. Пухова НАН України. – Київ, 2017. – Вип. 79. – С.3-8.

24. Гильгурт С.Я. Анализ применения аппаратного ускорения информационной защиты в автоматизированных системах энергетической отрасли / С.Я. Гильгурт // Моделювання та інформаційні технології. Зб. наук. пр. ІПМЕ ім. Г.Є. Пухова НАН України. – Київ, 2018. – Вип. 83. – С.154-164.

25. Евдокимов В.Ф. Дополнительные этапы процедуры оперативной

реконфигурации аппаратных ускорителей задач информационной безопасности / В.Ф. Евдокимов, А.Н. Давиденко, С.Я. Гильгурт // Моделювання та інформаційні технології. Зб. наук. пр. ІПМЕ ім. Г.Є. Пухова НАН України. – Київ, 2018. – Вип. 85. – С.3-11.

26. Евдокимов В. Централизованный синтез реконфигурируемых аппаратных средств информационной безопасности на высокопроизводительных платформах / В. Евдокимов, А. Давиденко, С. Гильгурт // Захист інформації. – 2018. – Т. 20, № 4. – С.247-258.

27. Гільгурт С.Я. Побудова асоціативної пам'яті на цифрових компараторах реконфігуровними засобами для вирішення задач інформаційної безпеки / С.Я. Гільгурт // Електронне моделювання. – 2019. – Т. 41, № 3. – С.59-80.

28. Гільгурт С. Побудова фільтрів Блума реконфігуровними засобами для вирішення задач інформаційної безпеки / С. Гільгурт // Безпека інформації. – 2019. – Т. 25, № 1. – С.53-58.

29. Гільгурт С. Побудова скінченних автоматів реконфігуровними засобами для вирішення задач інформаційної безпеки / С. Гільгурт // Захист інформації. – 2019. – Т. 21, № 2. – С.111-120.

30. Гільгурт С. Методи побудови оптимальних схем розпізнавання для реконфігуровних засобів інформаційної безпеки / С. Гільгурт // Безпека інформації. – 2019. – Т. 25, № 2. – С.74-81.

Праці апробаційного характеру:

31. Гильгурт С.Я. Применение типовых устройств на базе программируемой логики для решения вычислительных задач / С.Я. Гильгурт // Труды II международной конф. «Параллельные вычисления и задачи управления» PACO’2004 памяти Е.Г. Сухова. Москва, 4–6 окт. 2004 г. – М.: Институт проблем управления им. В.А. Трапезникова РАН, 2004. – С.514-530.

32. Гильгурт С.Я. К вопросу о применении реконфигурируемых вычислителей для решения ресурсоемких задач / С.Я. Гильгурт // Информационные технологии в управлении энергетическими системами (ИТУЭС-2005): тез. докл. Междунар. конф. (18 – 19 октября 2005 р., г. Киев), К.: Институт проблем моделирования в энергетике им. Г.Е. Пухова НАН Украины, 2005. – С.72-74.

33. Гильгурт С.Я. Решение задач распознавания с применением реконфигурируемых вычислителей / С.Я. Гильгурт, А.Н. Давиденко // Искусственный интеллект. Интеллектуальные и многопроцессорные системы – 2006: Тез. докл. Международной научно-технической конференции. Т. 2. – Таганрог: Изд-во ТРТУ, 2006. – С.76-79.

34. Гильгурт С.Я. Применение типовых реконфигурируемых устройств на базе ПЛИС для ресурсоемких вычислений / С.Я. Гильгурт // Матеріали Міжнар. наук.-практ. конф. «Інтелектуальні системи прийняття рішень та інформаційні технології», м. Чернівці, 17-19 травня 2006 р. – Чернівці: "Рута", 2006. – С.278-279.

35. Гильгурт С.Я. О применении реконфигурируемых унифицированных вычислителей для решения научно-технических задач / С.Я. Гильгурт // Параллельные вычислительные технологии (ПаВТ'2008): Труды международной научной конференции (Санкт-Петербург, 28 января – 1 февраля 2008 г.). –

Челябинск: Изд. ЮУрГУ, 2008. – С.358-363.

36. Гильгурт С.Я. Некоторые вопросы аппаратного ускорения в современных информационных технологиях / С.Я. Гильгурт // Материалы Международной научно-технической конференции «Многопроцессорные вычислительные и управляющие системы» (МВУС-2009), Дивноморское, 28 сентября – 03 октября 2009 г. Т. 1. – Таганрог: Изд-во ТТИ ЮФУ, 2009. – С.26-29.

37. Гильгурт С.Я. Анализ применения реконфигурируемых устройств в системах обнаружения вторжений / С.Я. Гильгурт // Тез. доп. Міжнар. наук.-техн. конф. «Моделювання-2010». Т. 1. – Київ: Інститут проблем моделювання в енергетиці ім. Г.Є. Пухова НАН України, 2010. – С.260-267.

38. Гильгурт С.Я. Аппаратное распознавание строк в интеллектуальных системах защиты информации / С.Я. Гильгурт // Тез. докл. международной научно-технической конференции «Искусственный интеллект. Интеллектуальные системы ИИ-2011». – Донецк: ИПИИ «Наука і освіта». – 2011. – Т. 1. – С.271-274.

39. Гільгурт С.Я. Програмно-апаратна система захисту даних для інфраструктури грід / С.Я. Гільгурт, А.К. Гіранова // Сучасні комп'ютерні системи та мережі: розробка та використання: матеріали 5-ої Міжнар. наук.-техн. конф. ACSN-2011, 29 вересня – 1 жовтня 2011, Львів, Україна. – Л.: Вид-во Нац. ун-ту «Львів. політехніка», 2011. – С.50-53.

40. Гільгурт С.Я. Множинне розпізнавання рядків у системах виявлення вторгнення на базі реконфігурованих обчислювачів / С.Я. Гільгурт // Сучасні комп'ютерні системи та мережі: розробка та використання: матеріали 5-ої Міжнар. наук.-техн. конф. ACSN-2011, 29 вересня – 1 жовтня 2011, Львів, Україна. – Л.: Вид-во Нац. ун-ту «Львів. політехніка», 2011. – С.54-56.

41. Гильгурт С.Я. Аппаратное ускорение задач информационной безопасности на базе ПЛИС с применением грид-вычислений / С.Я. Гильгурт // Сучасні проблеми і досягнення в галузі радіотехніки, телекомунікацій та інформаційних технологій: Тез. доп. VII Міжнар. наук.-прак. конф. (17–19 вересня 2014 р., м. Запоріжжя), Запоріжжя: ЗНТУ, 2014. – С.329-330.

42. Гильгурт С.Я. Задача множественного распознавания строк в интенсивном потоке данных и методы ее аппаратного ускорения / С.Я. Гильгурт // Тез. доп. Міжнар. наук.-техн. конф. «Моделювання-2016». – Київ: Інститут проблем моделювання в енергетиці ім. Г.Є. Пухова НАН України, 2016. – С.166-169.

43. Гильгурт С.Я. Применение реконфигурируемых вычислителей для аппаратного ускорения сигнатурных систем защиты информации / С.Я. Гильгурт // Тез. доп. Міжнар. наук.-техн. конф. «Моделювання-2018». – Київ: Інститут проблем моделювання в енергетиці ім. Г.Є. Пухова НАН України, 2018. – С.107-110.

44. Hilgurt S.Ya. A method to construct the signature-based secure tools on reconfigurable accelerators / S.Ya. Hilgurt // Сучасні проблеми і досягнення в галузі радіотехніки, телекомунікацій та інформаційних технологій: Тез. доп. IX Міжнар. наук.-прак. конф. (03–05 жовтня 2018 р., м. Запоріжжя) [Електронний ресурс] / Редкол.: Д.М. Піза, С.В. Морщавка. Електрон. дані. – Запоріжжя: ЗНТУ, 2018. – 1 електрон. опт. диск (DVD-ROM); 12 см. – Назва з тит. екрана, 2018. – С.200-201.

45. Гильгурт С.Я. Централизованный синтез в грид-среде реконфигурируемых средств защиты информации / С.Я. Гильгурт // Тез. доп. Міжнар. наук.-техн. конф.

«Високопродуктивні обчислення» (HPC-UA 2018) – Київ, 2018. – С.40-47.

46. Гильгурт С.Я. Применение реконфигурируемых устройств на базе ПЛИС при построении сетевых систем обнаружения вторжений / С.Я. Гильгурт // Перспективні напрями захисту інформації: матеріали IV всеукр. наук.-пр. конф. – м. Одеса, 02-06 вересня 2018 р.: збірник тез. – Одеса: ОНАЗ ім. О.С. Попова, 2018. – С.37-42.

47. Гільгурт С.Я. Підвищення ефективності реконфігуровних систем виявлення вторгнень / С.Я. Гільгурт // Безпека інформаційних технологій: матеріали IX Міжнар. наук.-техн. конф. ITSec-2019, 22-27 березня 2019, м. Шарм-ель-Шейх, Єгипет. – К.: НАУ, 2019. – С.10-11.

48. Давыденко А.Н. Применение грид-сети для синтеза промышленных систем защиты информации на базе ПЛИС / А.Н. Давыденко, С.Я. Гильгурт // «Цифровые технологии в промышленности»: материалы республиканской научно-практической конференции, г. Актау, Казахстан, 28 марта 2019 г. – Актау, КГУТИ им. Ш. Есенова, 2019. – С.15-20.

49. Hilgurt S. Method for constructing reconfigurable multi-pattern matching modules for information security systems / S Hilgurt // Захист інформації і безпека інформаційних систем: матеріали VII Міжнар. наук.-техн. конф, м. Львів, 30 – 31 травня 2019. – Львів: Видавництво Львівської політехніки, 2019. – С.134-135.

50. Гільгурт С.Я. Апаратне рішення задач кібербезпеки в електроенергетичній галузі / С.Я. Гільгурт // Кібербезпека енергетики: Збірка праць конференції, м. Одеса, 28 травня – 1 червня 2019. – К.: Інститут проблем моделювання в енергетиці ім. Г.Є. Пухова НАН України, 2019. – С.11-14.

51. Давиденко А.М. Методи підвищення ефективності мережевих систем виявлення вторгнень на базі ПЛІС / А.М. Давиденко, С.Я. Гільгурт, О.О. Політучий // Матеріали V Всеукр. наук.-практич. конф. «Перспективні напрями захисту інформації 2019», смт Затока Одеської обл., 31 серпня – 5 вересня 2019. – Одеса: 2019. – С.52-54.

Праці, які додатково відображають наукові результати дисертації:

52. Патент UA 139730 U; G06F17/27; Апаратно-програмний комплекс підтримки прийняття рішень при проведенні державних експертиз комплексних систем захисту інформації / Давиденко А.М., Гільгурт С.Я., Шабан М.Р.; Інститут проблем моделювання в енергетиці ім. Г.Є. Пухова Національної академії наук України. – заяв. u 2019 09353, 16.08.2019 р. – Опубл. 10.01.2020, Бюл. № 1.

АНОТАЦІЯ

Гільгурт С.Я. Методи та засоби створення реконфігуровних сигнатурних засобів захисту інформації комп'ютерних систем і мереж. – На правах рукопису.

Дисертація на здобуття наукового ступеня доктора технічних наук за спеціальністю 05.13.05 – комп'ютерні системи та компоненти. – Інститут проблем моделювання в енергетиці ім. Г.Є. Пухова НАН України, Київ, 2020.

Дисертаційна робота присвячена розробці та розвитку методів побудови ефективних реконфігуровних апаратних сигнатурних засобів технічного захисту

інформації шляхом створення комбінованих обчислювальних структур, оптимізованих за заданими показниками в заданих умовах.

Для досягнення мети були проаналізовані як проблеми технічного захисту інформації, так і можливості сучасних ПЛІС для їх вирішення. Сформульовано та класифіковано показники ефективності розробок сигнатурних засобів захисту як технічних систем. В результаті вивчення світового досвіду було виокремлено три найбільш ефективних підходи до побудови апаратних засобів інформаційного захисту. Шляхом аналізу та систематизації відомих підходів та їх модифікацій виявлені та формалізовані властивості кожного з них, враховуючи специфіку реалізації на ПЛІС.

Для підвищення ефективності створюваних пристроїв розроблено та вдосконалено методи комбінування в одному модулі декількох різнорідних підходів, які шляхом виконання процедур оптимізації дозволяють максимізувати переваги кожного з них. Розроблено методи прискореного обчислення кількісних значень технічних характеристик компонентів реконфігуровних засобів захисту без виконання повного циклу часоємних і ресурсоємних процедур синтезу обчислювальної структури та конфігурацій для завантаження в ПЛІС. Розроблено алгоритми реалізації запропонованих методів комбінування та програмні засоби для їх перевірки і аналізу.

З метою спрощення практичного застосування запропонованих методів і засобів був реалізований у вигляді макетного зразка веб-сервіс централізованого створення реконфігуровних прискорювачів з використанням розподілених і хмарних обчислювальних середовищ. Сервіс дозволяє користувачам, які не мають навичок самостійної розробки обчислювальних структур для ПЛІС і програмування реконфігуровних пристроїв, завантажувати в систему перелік сигнатур, параметри наявних прискорювачів і додаткові вимоги до створюваних засобів технічного захисту інформації, віддалено спостерігати за ходом процесу їх синтезу та отримувати файли конфігурації для ПЛІС своїх систем інформаційного захисту.

Ключові слова: технічний захист інформації, сигнатура, патерн, множинне розпізнавання рядків, ПЛІС, реконфігуровний обчислювач, ефективність, комбінування.

ABSTRACT

Hilgurt S.Ya. Methods and techniques for creating reconfigurable signature-based security tools for computer systems and networks. – As the manuscript.

Thesis for a Doctor of Technical Sciences in Specialty 05.13.05 – Computer Systems and Components. – Pukhov Institute for Modelling in Energy Engineering, National Academy of Sciences of Ukraine, Kyiv, 2020.

The thesis is devoted to the research and development of methods for constructing effective reconfigurable hardware signature-based means of information security by combining computing structures, optimized according to specified criteria under specified conditions.

To achieve the goal, both the problems of technical protection of information and the capabilities of modern FPGAs as a hardware platform were analyzed. Criteria of the efficiency of signature protection means as technical systems have been formulated and

classified. As a result of studying the world experience, the three most effective approaches to the construction of hardware information protection systems were identified. By analyzing and systematizing the known approaches, as well as their modifications, the key features of each of them have been identified and formalized, taking into account the questions of the implementation on FPGA.

To increase the efficiency of the devices being created, methods for combining several different approaches in one module have been developed, which, by performing optimization procedures, maximize the benefits of each of them. Methods have been developed for the accelerated calculation of quantitative values of the technical characteristics of the components of reconfigurable protection means used as target functions for optimization procedures, without performing a full cycle of time-consuming and resource-intensive procedures for synthesizing a computational structure and bitstreams for loading into an FPGA. Algorithms for the implementation of the proposed combining methods have been developed, software tools have been made for their verification and analysis.

In order to simplify the practical application of the proposed methods and tools, a web service using distributed and cloud computing was implemented as a prototype for centralized creation of reconfigurable accelerators. The service allows users who do not have the skills to configure and programming reconfigurable devices, to upload a list of signatures, parameters of available accelerators and additional requirements into the system. Then they can remotely monitor the progress of the synthesis and finally receive the generated bitstreams for the FPGAs of their information security systems.

Keywords: information security, signature, DPI, multi-pattern string matching, FPGA, reconfigurable accelerator, efficiency, approach combining.

АННОТАЦИЯ

Гильгурт С.Я. Методы и средства создания реконфигурируемых сигнатурных средств защиты информации компьютерных систем и сетей. – На правах рукописи.

Диссертация на соискание ученой степени доктора технических наук по специальности 05.13.05 – компьютерные системы и компоненты. – Институт проблем моделирования в энергетике им. Г.Е. Пухова НАН Украины, Киев, 2020.

Диссертация посвящена разработке и развитию методов построения эффективных реконфигурируемых аппаратных сигнатурных средств технической защиты информации путем создания комбинированных вычислительных структур, оптимизированных по заданным показателям в заданных условиях.

Для достижения цели были проанализированы как проблемы технической защиты информации, так и возможности современных ПЛИС в качестве аппаратной базы для их решения.

В результате изучения многочисленных разработок получена обобщенная структура сигнатурного средства информационной защиты, реализованной на ПЛИС. Выяснено, что ключевым компонентом такой системы является модуль распознавания, который выполняет наиболее трудоемкую функцию – решает задачу множественного распознавания строк. Как следствие, от свойств данного модуля в значительной степени зависят результирующие характеристики всей системы

защиты в целом. Специфической особенностью применения реконфигурируемых средств для аппаратного ускорения работы сигнатурных систем защиты является регулярно возникающая необходимость повторного синтеза некоторых компонентов. Такая операция нужна, во-первых, при появлении описаний новых атак, во-вторых – при изменении условий работы защищаемой информационной системы.

Чтобы иметь возможность оценивать и сравнивать свойства разработок сигнатурных средств защиты и их компонентов, сформулированы и классифицированы показатели их эффективности.

В результате анализа имеющегося мирового опыта было выделено три наиболее эффективных подхода к построению аппаратных средств информационной защиты: ассоциативная память на базе цифровых компараторов; фильтр Блума на базе хеш-функций; алгоритм Ахо-Корасик на базе конечных автоматов. Изучены и формализованы свойства каждого из них с учетом особенностей реализации на реконфигурируемых средствах. Выяснилось, что ни один из исследованных подходов не обладает явным превосходством по сравнению с другими. Каждому свойственны преимущества и недостатки.

В этой связи естественным образом возникает идея объединить в одном модуле нескольких разнородных подходов таким образом, чтобы максимизировать преимущества каждого из них. Для ее реализации были разработаны и усовершенствованы методы комбинирования, позволяющие добиться повышения эффективности результирующего решения за счет применения оптимизационных процедур.

Разработаны методы ускоренного вычисления количественных значений технических характеристик компонентов реконфигурируемых средств защиты, позволяющие избавиться от времяемких и ресурсоемких процедур полного синтеза вычислительной структуры для ПЛИС. Суть методов заключается в формировании так называемых функций оценки, которые представляют собой функциональные зависимости технических характеристик синтезируемых схем распознавания от свойств набора сигнатур, подлежащих распознаванию с одной стороны, и параметров реконфигурируемого ускорителя – с другой. В диссертации разработаны алгоритмы реализации предложенных методов и программные средства для их проверки и анализа.

Как показали проведенные эксперименты, параллельное комбинирование различных подходов и их модификаций для аппаратного распознавания сигнатур позволяет улучшить показатели итоговой схемы в отдельных случаях на 35-40%, последовательное каскадирование – на 40-45%. Рассмотрен также принцип комбинирования методов комбинирования, позволяющий компоновать изученные методы в иерархичную структуру произвольной глубины.

С целью упрощения практического применения предложенных методов и средств был реализован в виде макетного образца веб-сервис централизованного создания реконфигурируемых ускорителей с использованием распределенных и облачных вычислительных сред. Сервис позволяет пользователям, которые не имеют навыков самостоятельной разработки вычислительных структур для ПЛИС и программирования реконфигурируемых устройств, загружать в систему перечень

сигнатур, параметры имеющихся ускорителей и дополнительные требования к создаваемым средствам технической защиты информации, в удаленном режиме наблюдать за ходом процесса их синтеза и получать созданные файлы конфигурации для ПЛИС своих систем информационной защиты.

Ключевые слова: техническая защита информации, сигнатура, задача множественного распознавания строк, ПЛИС, реконфигурируемы вычислитель, повышение эффективности, комбинирование.

Підписано до друку 09.11.2020 р. Формат 60х90 1/16.
Папір офсетний. Умовн. др. арк. 1,9.
Наклад 100 прим. Зам. № 0911/01.

Надруковано ФОП Гузік О.М.
Реєстраційний номер №2705814113
м. Київ, вул. Б. Гаврилишина, 16
Тел.: 338-16-61.