УДК 004.274:004.056

В.Ф.Євдокимов, А.М.Давиденко, С.Я.Гільгурт, м. Київ

СТВОРЕННЯ НА БАЗІ ГРІД-САЙТУ ІПМЕ ІМ. Г.Є. ПУХОВА НАНУ СИСТЕМИ ЦЕНТРАЛІЗОВАНОГО СИНТЕЗУ АПАРАТНИХ ПРИСКОРЮВАЧІВ ДЛЯ ВИРІШЕННЯ ЗАДАЧ ІНФОРМАЦІЙНОЇ БЕЗПЕКИ В ЕНЕРГЕТИЧНІЙ ГАЛУЗІ[1]

Abstract. Methods to synthesize reconfigurable accelerators aimed to solve computer security tasks are investigated. Using GRID infrastructure as a platform for organizing a remote centralized service for this purpose is proposed. Grid-node of Pukhov Institute for Modelling in Energy Engineering is regarded as a central unit of such system.

Вступ

Програмна реалізація складних інструментів інформаційної безпеки, таких як системи виявлення вторгнень (СВВ) та антивірусні засоби, стає щодня більш проблематичною у зв'язку зі збільшенням числа і складності комп'ютерних атак, а також через припинення зростання частоти універсальних мікропроцесорів. Ці чинники змушують розробників звертатися до рішень на базі програмованих логічних інтегральних схем (ПЛІС). Але процес розробки та конфігурування апаратного пристрою на базі ПЛІС є складною та ресурсномісткою задачею. Користувачі систем інформаційної безпеки (системні адміністратори) не мають можливості для самостійної розробки апаратних антивірусів та СВВ.

Тому набувають актуальність рішення, за якими процес синтезу реконфігуровних засобів організований таким чином, що складні та ресурсномісткі процедури виконуються не локально на кожної окремої системі, а централізовано, з використанням високопродуктивних систем, зокрема, грід-інфраструктури.

Аналіз останніх досліджень та публікацій свідчить, що питаннями використання реконфігуровних обчислювачів при створенні систем інформаційної безпеки займаються фахівці у багатьох країнах світу – США, Європі (зокрема в Швейцарії та Греції) і Південно-східної Азії (Індії, Японії, Китаю та ін.) [1 – 7]. Друкується велика кількість публікацій стосовно використання ПЛІС (типу FPGA) для прискорення ресурсномісткої задачі розпізнавання сигнатур в інтенсивному потоці даних. Але відомостей щодо робіт з використання розподілених обчислень, зокрема грід-технологій для централізованого синтезу цифрових структур, що завантажуються в ПЛІС на

[1] Исследование выполнено при частичном финансировании Целевой комплексной программой научных исследований НАН Украины «Грид-инфраструктура и грид-технологии для научных и научно-прикладных применений», 2016 г.

сьогодні немає.

У країнах СНД існує низка наукових колективів та шкіл, де займаються проблемами використання ПЛІС та реконфігуровними обчисленнями, включаючи задачі інформаційної безпеки, але не питаннями використання апаратних прискорювачів для систем виявлення вторгнень або антивірусних систем.

Метою даної роботи є дослідження та розробка принципів побудови систем централізованого синтезу конфігурацій апаратних прискорювачів на базі ПЛІС для вирішення задач інформаційної безпеки в енергетичній галузі з використанням досвіду працівників Інституту проблем моделювання в енергетиці ім. Г.Є. Пухова НАН України.

1. Досвід колективу ІПМЕ ім. Г.Є. Пухова НАНУ

Дослідженнями в галузі грід-технологій працівники Інституті проблем моделювання ім. Г.Є. Пухова НАН України займаються з 2007 р. В 2008 р побудовано і підключено до національної грід-мережі обчислювальний кластер Інституті проблем моделювання ім. Г.Є. Пухова НАН України (PIMEE). В 2012 р створена віртуальна організація з математичного моделювання в задачах енергетики MatModEn. В 2012 р грід-сайт зареєстровано (з ім'ям UA-PIMEE) та сертифіковано у європейської грід-інфраструктурі EGI [8, 9].

Весь час з моменту запуску в жовтні 2008 р. грід-вузол PIMEE/UA-PIMEE працював у цілодобовому режимі (з перервами на модернізацію двічі по декілька діб). Тести на основі систем моніторингу УНГ та NGI_UA проходять вдало. Відповідно до даних системи моніторингу EGI усереднений коефіцієнт доступності грід-сайту за 2016 р. складав 0,96 та є одним з найбільших серед усіх грід-сайтів NGI_UA.

Фахівці ІПМЕ ім. Г.Є. Пухова НАНУ також на протязі багатьох років працюють над проблемами розробки високопродуктивних та гнучких засобів, у тому числі – реконфігуровних, яки будуються на базі ПЛІС [10 – 14].

Таким чином, авторський колектив, з одного боку, має великий досвід з побудови реконфігуровних засобів для високопродуктивної обробки даних, з іншого боку, успішно володіє сучасними технологіями розподілених обчислень.

2. Практичний доробок

У 2015-2016 роках в рамках цільової комплексної програми наукових досліджень НАН України «Грід-інфраструктура і грід-технології для наукових і науково-прикладних застосувань» за участю фахівців ІПМЕ ім. Г.Є. Пухова НАНУ був розроблений у вигляді макетного зразка сервіс централізованого синтезу конфігурацій для апаратних прискорювачів [15]. В основу розробки було покладено новітній підхід, розроблений в рамках ВО medgrid та ВО moldyngrid – запуск віртуальної машини як грід-задачі за допомогою системи Rainbow ("ARC in the Cloud") [16].

Ця розробка дозволила перевірити та підтвердити ефективність ідеї переносу ресурсномістких операцій з локальних систем виявлення вторгнень

до високопродуктивного обчислювальних середовища, такого як грід-система, з метою централізованого виконання.

3. Наукова проблема створення блоків розпізнавання реконфігуровних засобів інформаційної безпеки

3.1. Внутрішній паралелізм задачі розпізнавання. Модуль розпізнавання сигнатур є найбільш важливим компонентом системи виявлення вторгнень та антивірусних засобів, від успішної реалізації якого суттєво залежать показники її ефективності. Отже, вибір алгоритму розпізнавання і технічного рішення для його реалізації є ключовими моментами при створенні таких систем.

Проте, задачі розпізнавання даних у мережевому трафіку в значній мірі властивий паралелізм, причому, за двома напрямками: по-перше, декілька елементів інформації можуть аналізуватися одночасно; по-друге, порівняння може робитися відразу з багатьма записами бази даних сигнатур [17]. Розглянемо ці напрями можливого розпаралелювання.

На жаль, при реалізації паралелізму першого типу виникає протиріччя: розділення інтенсивного вхідного потоку на велике число окремих блоків, що обробляються незалежними обчислювальними модулями, призводить до затримок, пропорційних коефіцієнту розпаралелювання, обумовленим великим розміром блоків; зменшення ж розміру блоків знижує корисність розпаралелювання через ефект перекриття, тим більшого, чим довші шукані підрядки. До того ж, такий підхід вимагає реалізації складних процесів управління, планерування і буферизації [6].

Розпаралелювання за базою сигнатур, тобто, розділення набору розпізнаваних підрядків на підгрупи також можливе. Але в цьому випадку виявляється важлива особливість – велика кількість сигнатур багато в чому повторюють одна одну. Причому, даний ефект самоподібності через скінченність алфавіту, теоретично, повинен зростати у міру зростання баз даних сигнатур. Облік такого ефекту дозволяє істотно підвищити продуктивність підсистеми, що розпізнає. Проте, найбільш розповсюджені одношаблонні алгоритми непридатні для даної мети.

3.2. Задача множинного розпізнавання рядків. Таким чином, виникає теоретична проблема множинного розпізнавання рядків [6]. Її суть полягає в одночасному пошуку у вхідній послідовності символів не одного підрядка, а заданого набору підрядків, різні фрагменти яких повторюються в значній мірі. Відомі способи розпаралелювання не дозволяють досягти прийнятного результату через вказані вище причини. Отже, ефективне рішення даної задачі можна отримати лише на рівні алгоритму або обчислювальної структури.

На сьогоднішній день в МСВВ на базі реконфігуровних сопроцесорів успішно застосовуються такі підходи і технічні рішення, як [1 – 5]:

- цифрові автомати (ЦА);
- паралельні дискретні компаратори;

- пристрої асоціативної пам'яті та її різновиди;
- різні варіанти використання хеш-функцій, зокрема, фільтр Блума.

Кожен з напрямів має як деякі переваги перед іншими, так і недоліки. Так, цифрові автомати, синтезовані в ПЛІС, не забезпечують високу пропускну спроможність, складні в побудові та конфігуруванні. Паралельні компаратори за більшої продуктивності приводять до підвищених витрат устаткування і погано масштабуються. Рішення, що базуються на асоціативній пам'яті, менш вимогливі до ПЛІС, ніж цифрові компаратори при сумірній продуктивності, але дорожче і споживають більше енергії. Фільтр Блума і стискування бази сигнатур функціями хешування дозволяють зменшити кількість порівнянь, але забезпечують імовірнісне розпізнавання, що вимагає додаткових витрат на уточнення результатів збігу.

3.3. Вимоги до апаратних прискорювачів. Таким чином, вибір найбільш ефективного підходу є досить складною задачею. Для її вирішення необхідно сформулювати вимоги, що пред'являються до апаратних прискорювачів на базі ПЛІС, а також основні параметри для оцінювання їх ефективності.

Головними показниками продуктивності МСВВ є: *максимальне число рядків*, що розпізнаються системою, та *пропускна здатність*, яка може при цьому бути досягнута [14]. Важливою характеристикою МСВВ, також пов'язаною з продуктивністю, є *передбачуваність пропускної здатності*, тобто, незалежність її часових характеристик від складу вхідних даних. Виявлення зловмисного контенту у мережному трафіку є рідкісним випадком, ймовірність виникнення якого в штатному режимі замала. Проте, якщо вміст аналізованих мережних пакетів істотно впливає на швидкодію модуля розпізнавання МСВВ, то така система може виявитися уразливою до навмисного засмічення мережного трафіку сигнатурами відомих атак зловмисником. Крім швидкісних характеристик систем виявлення вторгнення, для їх практичного використання важливі також вартісні показники. *Об'єм оперативної* пам'яті, необхідної для реалізації вибраного алгоритму розпізнавання істотно впливає, в результаті, на швидкодію. Якщо ресурсів швидкодіючої блокової пам'яті (BRAM), що має ПЛІС, недостатньо для реалізації запам'ятовуючого пристрою, то виникає необхідність у зовнішній пам'яті, яка набагато повільніша внутрішньої. *Незалежність від складу сигнатур* також є важливою характеристикою МСВВ. Орієнтація модуля розпізнавання на обмежений алфавіт з метою підвищення швидкодії може призвести до небажаних наслідків при його використанні в МСВВ. Специфічною рисою систем виявлення вторгнення на базі сигнатур є необхідність регулярного оновлення активної бази даних. Можливості *динамічного оновлення* істотно впливають на практичну корисність технічного рішення. Істотною є також *загальна вартість реалізації* системи. Яким би ефективним не був модуль розпізнавання, якщо для його інтеграції в МСВВ необхідні істотні додаткові витрати, наприклад, на перетворення форми представлення інформації, загальна вартість рішення може виявитися незадовільною.

3.4. Рекомендоване рішення. В якості платформи для створення блоків розпізнавання реконфігуровних засобів інформаційної безпеки за результатами аналізу існуючих розробок за переліченими показниками з невеликим відривом можна віддати перевагу рішенню, запропонованому в роботі [6]. Його основою є цифрові автомати Ахо-Корасік. Висока швидкодія досягається за рахунок конвеєрної обробки інформації на двох рівнях: всередині самих ЦА, а також на загально-структурному рівні (рис. 1).

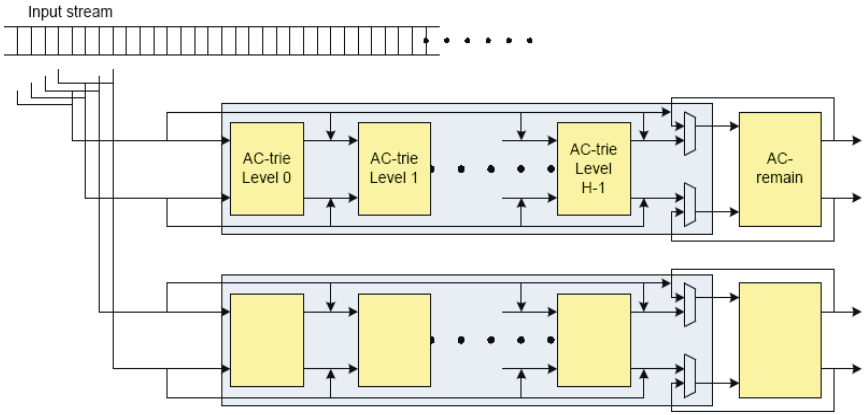

Рис. 1. Мультіконвеєрна структура високопродуктивного апаратного розпізнавання сигнатур в інтенсивному вхідному потоці байтів

За рахунок цього вдається ефективно використовувати властивий паралелізм задачі розпізнавання рядків, що вирішується модулем розпізнавання сигнатур системи виявлення вторгнень.

Висновки

В даній роботі розглянута проблема використання реконфігуровних обчислювальних засобів для створення систем інформаційної безпеки.

Досліджені питання побудови на базі ІПМЕ ім. Г.Є. Пухова НАНУ компонентів систем централізованого синтезу апаратних прискорювачів для вирішення задач інформаційної безпеки в енергетичній галузі.

Проведений аналіз світового досвіду в галузі розробок засобів інформаційної безпеки на базі реконфігуровних обчислювачів. Сформульовані критерії вибору, запропоновано рішення.

1. *Sidhu R., Prasanna V.* Fast regular expression matching using FPGAs // IEEE Symposium on Field Programmable Custom Computing Machines (FCCM01), April 2001.
2. *Carver D., Franklin R., Hutchings B.* Assisting network intrusion detection with reconfigurable hardware // IEEE Symposium on Field-Programmable Custom Computing Machines (FCCM02), April 2002.
3. *Sourdis I., Pnevmatikatos D.* Pre-decoded CAMs for Efficient and High-Speed NIDS Pattern Matching // Proceedings of 12th IEEE Symposium on Field Programmable Custom

Computing Machines (FCCM04), April 2004.
4. *Gokhale M., Dubois D., Dubois A., Boorman M. etc.* Granidt: Towards gigabit rate network intrusion detection technology // Proceedings of the 12th International Conference on Field-Programmable Logic and Applications, Sept. 2002.
5. Sourdis I., Pnevmatikatos D. Fast, large-scale string match for a network intrusion detection system // Proceedings of 13th International Conference on Field Programmable Logic and Applications, 2003.
6. *Jiang W., Prasanna V.* Scalable Multi-Pipeline Architecture for High Performance Multi-Pattern String Matching // IEEE International Parallel and Distributed Processing Symposium (IPDPS '10), April 2010.
7. *Cho Y., Navab S., Mangione-Smith W.* Specialized Hardware for Deep Network Packet Filtering // Proceedings12th International Conference on Field-Programmable Logic and Applications, Sept. 2002.
8. *Евдокимов В.Ф., Давиденко А.Н., Гильгурт С.Я. та ін.* Грид-центр для решения задач моделирования / Моделирование и компьютерная графика // Материалы Третьей международной научно-технической конференции. Донецк: ДонНТУ, 2009. С. 8–12.
9. *Давиденко А.Н., Гильгурт С.Я., Душеба В.В., Гиранова А.К.* Технические средства дополнительной защиты данных пользователей в распределенных информационных системах // III Міжнародна науково-технічна конференція «Комп'ютерні системи та мережні технології» / Тези доп., 15-17 червня 2010 р. – К.: Вид-во Нац.авіац.ун-ту «НАУ-друк», 2010. – С. 32.
10. *Гильгурт С.Я.* Реконфигурируемые вычислители. Аналитический обзор // Электронное моделирование. – 2013. – Т.35, № 4. – С. 49–72.
11. *Гильгурт С.Я., Гиранова А.К.* Методика создания реконфигурируемых процессоров, реализующих усиленные алгоритмы закрытия информации // Зб. наук. пр. ІПМЕ НАН України. – Київ, 2011. – Вип. 61. – С. 69–78.
12. *Hilhurt S. Ya.* Application of FPGA-based Reconfigurable Accelerators for Network Security Tasks // Collection of scientific works. Simulation and informational technologies. – PIMEE NAS of Ukraine. – Kyiv, 2014. – Vol. 73. – P. 17–26.
13. *Давиденко А.Н, Гильгурт С.Я.* Алгоритмы распознавания строк в системах обнаружения вторжений на ПЛИС // Моделювання та інформаційні технології. Зб. наук. пр. ІПМЕ НАН України. – Київ, 2010. – Вип. 58. – С. 103–109.
14. *Гильгурт С.Я.* Множинне розпізнавання рядків у системах виявлення вторгнення на базі реконфігуровних обчислювачів // Сучасні комп'ютерні системи та мережі: розробка та використання : матеріали 5-ої Міжнар. наук.-техн. конф. ACSN-2011, 29 вересня – 01 жовтня 2011, Львів, Україна. – Л. : Вид-во Нац. ун-ту «Львів. політехніка», 2011. – С. 54–56
15. *Гильгурт С.Я.* Организации вычислительного процесса синтеза файлов конфигураций для аппаратных ускорителей при решении задач информационной безопасности // Моделювання та інформаційні технології. Зб. наук. пр. ІПМЕ НАН України. – Київ, 2015. – Вип. 74.
16. *Сальников А.А., Вишневский В.В., Борецкий А.Ф.* «Платформа як сервіс» у грід для інтерактивного аналізу медичних даних // Математичні машини і системи. – 2015. – № 1. – С. 53–64.
17. *Давиденко А.Н., Гильгурт С.Я., Сабат В.И.* Аппаратное ускорение алгоритмов сигнатурного обнаружения вторжений в открытой системе информационной безопасности Snort // Моделювання та інформаційні технології. Зб. наук. пр. ІПМЕ НАН України. – Київ, 2012. – Вип. 65. – С. 94–103.

Поступила 3.04.2017р.