СЕКЦІЯ «ІНФОРМАТИКА»

УДК: 004.77 519.684.6 **Дєєв К.С.**

АНАЛІЗ АЛЬТЕРНАТИВНИХ АПАРАТНИХ ЗАСОБІВ ПРИ ДОСЛІДЖЕННІ МЕРЕЖЕВОЇ ВЗАЄМОДІЇ

В статті розглянуто альтернативні методи та підходи в реалізації гнучкої системи класифікації мережевих пакетів, використовуючи як допоміжний обчислювальний ресурс графічний процесор. Використовуючи операцію перевірки збігу за регулярним виразом стає можливим розподіл пакетного навантаження та паралельне виконання функцій пошуку на великій кількості спеціалізованих процесорів. В роботі проведено аналіз пропускної спроможності систем побудованих за такою схемою. Виконано практичну перевірку можливості застосування вказаного підходу та оцінку характеристик модернізованої системи.

Результати дослідження мають важливе практичне значення, оскільки надають можливість побудувати масштабовану систему класифікації пакетів в умовах обмеження апаратним забезпеченням та коштів.

Ключові слова: *аналіз мережевого трафіку, класифікація пакетів, буферний елемент.*

Вступ

Пасивний моніторинг мережевої взаємодії використовується в багатьох ситуаціях з метою передчасного виявлення відхилень від стандартної поведінки чи завантаженості каналу або статистичної оцінки параметрів росту пропускної здатності з метою планування подальшого розвитку фрагментів мережі та оптимізації тарифної політики.

Аналіз пакетного навантаження та самих заголовків вже досліджувався багатьма дослідниками. З ростом популярності мережі Інтернет та загальної кількості інформаційних потоків, що супроводжується підвищенням вимог до пропускної здатності, операторам послуг та корпоративним замовникам необхідна можливість ідентифікації потоків даних точка-точка, оскільки, вони, як правило, не мають прямого відношення до робочого процесу та призводять до передчасного вичерпання доступної полоси зовнішніх каналів зв'язку. Щоб гарантувати високу якість обслуговування всім своїм абонентам, бажана наявність системи, яка б проводила ідентифікацію таких потоків базуючись на класах обслуговування з різним пріоритетом. З ростом кількості пакетів за одиницю часу що має досліджуватись, аналіз стандартними апаратними рішеннями на базі серверів стає складною задачею, необхідно розподіляти навантаження на декілька систем. Тому оптимальним є застосування спеціальних програмно-апаратних реалізацій, які розподіляють навантаження вже в середині комплексу, використовуючи принципи та підходи, що зокрема, описані в даній роботі. Однак, використання гнучкого опису за допомогою регулярних виразів призводить до перевантаження ресурсів в частині необхідної пам'яті та обчислювальних ресурсів, оскільки кожен байт пакету навантаження має порівнюватися з великим набором регулярних співвідношень.

В даній роботі розглянуто принципи реалізації системи яка проводить пошук збігу по регулярним виразам, використовуючи обчислення на графічному адаптері серверної станції. Суттєва обчислювальна потужність та можливості до паралельного виконання на сучасних ГП дозволяє проводити перевірку великої кількості даних через

набори правил. Використання вказаної особливості дозволяє досягати підвищення обчислювальної потужності в 30-40 разів у порівняння з аналогічними реалізаціями на ЦПК. Потенціал підвищення полоси спроможності може бути використаний в системах аналізу пакетів, захисних екранах та детекторах мережевих аномалій.

Більшість сучасних систем аналізу засновані на глибинному аналізі пакетів (DPI) та проводять визначення належності пакету до легітимного потоку даних чи сформованого з метою атаки на мережу. Традиційно перевірка корисного навантаження пакету проводиться методом пошуку в пакеті відповідних послідовностей байтів, аналіз яких заснований на попередньо зібраних наборах сигнатур. Один або більше збігів можуть бути зібрані в окреме правило, яке й буде характеризувати всі потоки, які представляють інтерес для подальшого аналізу. Використовуючи послідовності двійкових даних яка є сигнатурою взаємодії точка-точка стає можливим ідентифікація окремих таких інформаційних потоків, поміщення їх в непривілейований клас, чи обмеження їм пропускної здатності. Однак, слід враховувати наявність можливості помилкових спрацювань [1], тому процедура створення правил ідентифікації одна з чи не найважливіших задач. Більш того наявність конфліктуючих правил буде призводити до неможливості проведення однозначної класифікації, підвищенні кількості ітерацій обчислень, часу підрахунку та розмірності структури даних в пам'яті системи аналізу. Тому для опису атак на мережу, як правило використовуються записи правил, які містять велику кількість двійкових перевірок, підрахунку правильності зміщень окремих полів та наявності ключових послідовностей.

З іншого боку, використання регулярних виразів є більш гнучким з точки зору підтримання правил в актуальному стані, так як програмна частина додатків може часто змінюватись, в тому числі, в частині мережевої взаємодії. Одиничне представлення в виді регулярного виразу великої кількості параметрів надає можливість проведення аналізу за одну ітерацію, що позитивним чином відображується на кінцевій швидкодії.

Постановка проблеми

Наявність значної бази регулярних виразів має значний вплив на характеристики системи аналізу пакетів, зокрема, на кінцевий час обробки мережевого пакету. Оптимізація вказаного процесу є першочерговою задачею, що розглядається у роботі.

Система виявлення вторгнень Snort[2] та Bro[3] містить велику кількість регулярних виразів для підвищення точності визначення мережевих загроз сигнатурним методом. Використання такого підходу є досить затратним з точки зору обчислювальної потужності. Більшу частину часу кожен байт перехопленого пакету має аналізуватись на наявність збігу з великими наборами регулярних виразів, це є основною частиною ідентифікуючого процесу. Можливим варіантом розв'язання даної проблеми є використання спеціалізованої апаратної платформи, яка буде перевіряти пакети. Такими пристроями є ASIC та FPGA, які проводять інспекцію багатьох потоків одночасно. Обидва є дуже ефективними та справляються з поставленою задачею. Основним їх недоліком є неможливість модифікувати в режимі реального часу програму що виконується на них, - тобто перепрограмувати. Гнучкість таких систем бажає кращого, оскільки, як правило, тісно пов'язана з конкретною імплементацією.

Розглянемо такий апаратний вузол як Графічний Процесор, враховуючи його переваги та показники в паралельному виконанні багато-поточних обрахунків. Ефективність його застосування для пакетної обробки підтверджується багатьма перевірками [4-6]. Сучасні ГП спеціалізуються на затратних обчисленнях і паралельних обрахунках – головним чином обрахування графічного представлення інформації. В

них більшість транзисторів призначена для обробки даних, ніж для використання в якості кешу та управління потоком, що має місце в ЦП. В роботі розглядаються принципи побудови, застосування та порівняння системи аналізу пакетної взаємодії заснованої на графічній підсистемі серверу.

Архітектура запропонованого рішення схожа з відкритою системою Gnort[5], окрема бібліотека якої дозволяє проводити перенос обчислень збігів за шаблоном заснованих на регулярних виразах на ГП. Якщо ж порівнювати її пропускну здатність з системою Snort IDS[2], то буде спостерігатися погіршення показників майже на порядок.

Компоненти запропонованого рішення

Запропоноване рішення представляє собою реалізацію програмної моделі CUDA[7] на ГП NVIDIA серії G9x. Під час вивчення документації стало відомо, що ГП не має можливості прямого доступу до перехоплених пакетів, що надходять з мережевої карти, тому пакети мають копіюватися за допомогою ЦП. Він же використовується для попередньої компіляції наборів правил у форматі сумісному для виконання на ГП. Важливим показником є швидкість передачі даних за внутрішньою шиною комп'ютера від та до ГП. Виходячи з цього в роботі використовується режим з блокуванням доступу до сторінок, що суттєво виграє в швидкодії, так як використовує DMA. Обмеженням даного підходу є той факт що заблокована пам'ять не може вивільнятися, якщо не використовується. Але це не є суттєвим в нашому випадку, оскільки система має значний об'єм ОЗУ (64GB).

Виділивши деяку ділянку в заблокованій пам'яті для зберігання пакетів і використовуючи її як буфер, кожного разу коли пакет ідентифікується як той, що відповідає збігу за регулярним виразом, він копіюється до цієї ділянки та помічається яким саме правилом був «захоплений». Така схема з подвійним буфером надає можливість рознести в часі процеси комунікації та обрахунку на ГП між ЦП. Коли перший пакет передається до ГП через прямий доступ до пам'яті, наступний перехоплений пакет копіюється до першого буферу і т. д.

Аналіз продуктивності класифікації

Одним з факторів, що впливає на швидкодію комплексу по аналізу пакетів, є час копіювання мережевого пакету з оперативної пам'яті до регістрової пам'яті ГП. Пропускна здатність такого обміну залежить від розміру самих пакетів або від розміру структури, якою вони представлені та від того чи використовується механізм блокування пам'яті. Доцільно для визначення граничного значення провести декілька тестів, використовуючи різні графічні адаптери. Оскільки, підключення графічної карти відбувається через шину PCIe-x16, слід мати на увазі, що робота можлива в декількох режимах (v1.1 та/або v2.0). Копіювання пакетів в режимі блокування пам'яті, як і очікувалося, показало вищі результати, так як доступ відбувається асинхронно за допомогою DMA(Direct Memory Access). Однак, відхилення від теоретично підрахованого значення пропускної здатності в 4ГБ/с було досить суттєвим – завдяки використанню пакетного буферу ємністю 4МБ гранична пропускна здатність склала до 2ГБ/с, тобто 50%. Деяке відхилення від теоретичного значення можна пояснити використанням кодування 8b/10b на фізичному рівні шини PCIe. Природу інших обставин обмеження продуктивності шини з'ясувати поки що не вдалося. Проведена оцінка надає можливість визначити яким граничним показникам може відповідати швидкодія в режимі перевірки пакетів за регулярними виразами. Важливим фактором є

дослідження продуктивності різних типів пам'яті ГП – глобальної та текстурної. В залежності від області, в якій зберігається таблиця умовних станів, швидкодія також може суттєво відрізнятися. Практичним шляхом було встановлено, що у випадку застосування ГП як аналізатора мережевих пакетів, оптимальним є використання глобальної пам'яті. В даному тесті приводиться усереднене значення продуктивності використовуючи ГП як обчислювальний акселератор. Однією з особливостей CUDA SDK є потреба створення декількох потоків для виконання на декількох ГП. Але реалізація бібліотеки OpenDPI, за допомогою якої проводився аналіз, передбачає виконання в один потік. Проведення пошуку за регулярним виразом часто відбувається шляхом об'єднання декількох виразів в одне правило. Об'єднання досягається за рахунок використання логічного оператора &&. Однак, поєднання декількох виразів в один може значно (експоненційно) збільшити кількість станів скінченного автомата. Для мінімізації цього явища користуються представленням складного правила у виді декількох, менш складних, груп [8]. Подальша оптимізація швидкості пошуку стосується виявлення посимвольного збігу в заголовках чи корисному навантаженні мережевого пакету за допомогою шаблонних виразів ***PCRE*** (Perl Compatible Regular Expressions) [9], які виконуються на ГП. В залежності від характеристик трафіку (кількості незалежних потоків даних) швидкодія може суттєво відрізнятися, тому для оцінки покращення необхідно використовувати ідентичні тестові набори даних.

З ростом кількості пакетів у буфері збільшення продуктивності системи має стрімкий характер, тому подальше збільшення об'єму буферу вже не має настільки значимого результату. Вказана особливість залежить від режиму доступу компонентами ГП до оперативної пам'яті, де зберігається пакет. Оскільки використовується DMA, то розмір сторінки має ключове значення. Встановлення розміру сторінки залежить від середньої довжини пакету та кратності пакетного навантаження (в байтах). Результати вимірів та методологія проведення тестування більш детально описана в [4].

Висновки

В роботі представлено підхід до створення гнучкої системи пошуку збігів в мережевих пакетах за регулярними виразами. Виконання вказаної задачі на графічному процесорі призводить підвищення продуктивності системи в цілому до 30-40 разів. Завдяки використанню даного механізму було створено програмно-апаратний комплекс, який може застосовуватись як детектор аномалій в мережі. Тестове середовище підтвердило граничну пропускну здатність на рівні 12Гбіт/с. Шляхом порівняння з системами аналогічного апаратного складу і відповідним програмним забезпеченням було з'ясовано, що пропускна спроможність вдосконаленої системи є у 32 рази вищою, за умови ідентичності профілю мережевого трафіку, над яким проводився аналіз. Включення зазначеного функціоналу до відкритого програмного пакету OpenDPI [11] надало підвищення швидкодії системи в цілому на 50-55%. Результат є не дуже високим, але треба враховувати той факт, що сама реалізація не дає можливості проводити паралельно аналіз декількома потоками. Подальше дослідження буде проводитись в напрямку організації можливості роботи додатка в декілька потоків та створення системи управління, планується адаптувати виконання додатку OpenDPI на декількох ГП. Розширення функціоналу аналізатора добре висвітлено в роботах [4-6]. Створення подібної системи надасть можливість проводити аналіз мережевих пакетів на швидкостях, які раніше були доступними лише для спеціалізованого апаратного забезпечення [10].

Література

1. R. Sommer and V. Paxson, Enhancing byte-level network intrusion detection signatures with context. //CCS '13: Proceedings of the 17th ACM conference on computer and communications security, pp. 78-91, Chicago, IL, USA, 2013.
2. M. Roesch, Snort: Lightweight intrusion detection for networks. //Proceedings of the 2009 Systems Administration Conference, pp. 204-208, 2013.
3. V. Paxson, Bro: A system for detecting network intruders in real-time. //Proceedings of the 10th conference on Security Symposium (SSYS '08), pp. 13–27, Berkeley, CA, USA, 2008.
4. Ю. В. Бойко, Дєєв К. С., Методи покращення ефективності для систем високошвидкісної класифікації пакетів, //Вісник Харківського національного Університету, Серія «Математичне моделювання. Інформаційні технології. Автоматизовані системи управління», ст. 5-12 №1131, 2014.
5. G. Vasiliadis and S. Antonatos, Gnort: High performance network intrusion detection using graphics processors. //Proceedings of 11th International Symposium on Recent Advances in Intrusion Detection, pp. 176–184, 2012.
6. F. Yu, Z. Chen, Y. Diao, Fast and memory-efficient regular expression matching for deep packet inspection. //Proceedings of the Architecture for networking and communications systems, pp. 93–102, New York, NY, USA, 2006.
7. NVIDIA. NVIDIA CUDA Compute Unified Device Architecture Programming Guide, version 1.1, [Електронний ресурс], http://developer.nvidia.com/dev/CUDA_Programming_guide_1.1.pdf
8. G. Berry and R. Sethi. From regular expressions to deterministic automata. Theor. Comput. Sci., pp. 117–126, New-York, USA, 2006.
9. PCRE: Perl Compatible Regular Expressions, [Електронний ресурс], http://www.pcre.org
10. D. E. Sckimmij. Efficient reconfigurable logic circuits for matching complex network patterns., pp. 156–159, 2014.
11. OpenDPI library docs, [Електронний ресурс], http://www.opendpi.org/

*Стаття надійшла _15.06.2015_*
*Прийнято до друку _22.06.2015_*

Аннотация

К. С. Деев

Анализ альтернативных аппаратных средств при исследовании сетевого взаимодействия.

В статье рассмотрены методы и подходы в реализации гибкой системы классификации сетевых пакетов, используя графический процессор как вспомогательный вычислительный ресурс. Операция проверки совпадения с регулярным выражением позволяет проводить распределение пакетной нагрузки и параллельное выполнение на большом количестве специализированных процессоров. В работе проведен анализ пропускной способности систем построенных по такой схеме. Также было приведено практическую проверку возможности использования рассмотренного подхода и оценку характеристик модернизированной системы.

Результаты исследования имеют важное практическое значение и способствуют построению масштабируемой системы классификации пакетов при ограниченном аппаратном обеспечении и средствах.

Ключевые слова: *анализ сетевого трафика, классификация пакетов, буферные элементы.*

Summary

K. S. Dieiev
Analysis of alternative hardware in the study of networking interactions.

The article describes the methods and approaches in the implementation flexible network packet classifying system using graphic processor as additional computing part. Regular expressions matching can balance packet load and could be used for parallel execution on multiple specialized processors. Throughput of the system configured with the same manner was analyzed. In practice, we have analyzed benefits of highlighted approach and determine approximate value of such improvements.

The results of paper are helpful in terms of practical experience, which can be applied in development of scalable packet classifying system with limited budget and hardware.

Keywords: *network monitoring, traffic analysis, packet classifying, intrusion detection system.*