УДК 004.725.5
Р.А. Бенца
Тернопільський національний технічний університет імені Івана Пулюя, Україна

АНАЛІЗ МЕРЕЖЕВОГО ТРАФІКА В РЕЖИМІ РЕАЛЬНОГО ЧАСУ

R.A. Bentsa

ANALYZE NETWORK TRAFFIC IN REAL TIME

Аналіз мережевого трафіку на сьогоднішній день - дуже велика тема. Під «Аналізом мережевого трафіку» ми будемо розуміти спільну назву технологій і їх реалізацій, що дозволяють проводити накопичення, обробку, класифікацію, контроль і модифікацію мережевих пакетів в залежності від їх вмісту в реальному часі. Одним з ускладнюючих чинників, при розгляді даного питання, є подвійність розвитку засобів аналізу мережевого трафіку: з одного боку - це розвиток алгоритмів і підходів до аналізу, з іншого - розвиток програмно-апаратних засобів для ефективного вирішення цього завдання. У свою чергу, це призводить як до плутанини в термінології, так і до свідомого маніпулювання фактами і цифрами в маркетингових цілях.

Зародження технологій аналізу мережевого трафіку можна віднести до початку 90-х років минулого століття. Потреби в їх виникненні з'явилися приблизно в водночас в декількох областях.

Ускладнення схем мереж і різноманіття мережевих пристроїв привели до ускладнення їх налаштування і підтримки мережі в працездатному стані - необхідний був інструмент який дозволяє, з одного боку локалізувати проблему, а з іншого надати якомога більш вичерпну інформацію про природу проблеми. Власне об'єктом, який містить в собі всю необхідну інформацію і є мережевий трафік. Одним з інструментів, спочатку призначеним для вирішення саме цієї проблеми став мережевий сніффер / аналізатор Wireshark, створений інженером Джеральдом Комбо (Gerald Comb) в 1997 році. Wireshark продовжує активно розвиватися і є стандартом в певній галузі мережевого аналізу.

Програмні і апаратні засоби, які здійснюють захоплення трафіку відносяться до класу сніффера (sniffers). Для вирішення завдання захоплення трафіку можуть використовуватися як стандартні серверні мережеві карти, так і спеціалізовані мережеві карти, призначені для перехоплення трафіку на граничних швидкостях без втрат. Спеціалізовані карти, як правило, реалізовані на базі FPGA або ASIC і мають вбудовані засоби для проставлення тимчасових міток, апаратної фільтрації, зняття деяких заголовків низькорівневих протоколів, балансування навантаження між процесорами на багатопроцесорних комп'ютерах з урахуванням IP-потоків, виявлення помилкових та дублюються пакетів. При цьому вся обробка (в тому числі і копіювання даних в пам'ять комп'ютера з пам'яті мережевої карти) здійснюється без залучення ресурсів ЦПУ. У міру розвитку технологій багато з описаних властивостей реалізуються і на базі стандартних мережевих карт.

Основною проблемою для стандартних мережевих адаптерів є не швидкість передачі даних, як така, а кількість пакетів в одиницю часу. Це обумовлено особливостями внутрішньої реалізації обробників пакетів на мережевих картах, драйверів мережевих карт і програмних мережевих стеків ОС. Внаслідок цього, стандартні мережеві карти без спеціалізованих драйверів і мережевих стеків не забезпечують перехоплення трафіку без істотних втрат на швидкостях понад 3 Mpps.

Більшість поширених систем аналізу трафіку працюють, використовуючи бібліотеки Libpcap (ОС Linux) і WinPcap (ОС Windows). Дані бібліотеки працюють в

режимі користувача. Для забезпечення своєї роботи з боку ОС вони використовують драйвери рівня ядра Berkeley Packet Filter (BPF) і Netgroup Packet Filter (NPF) відповідно. Основна різниця між цими драйверами полягає в схемі їх роботи з буферами пам'яті, які використовуються для тимчасового зберігання пакетів, одержуваних від мережевої карти. Драйвер BPF використовує схему з подвійною буферизацією, в той час як драйвер NPF використовує кільцевий буфер.

Для вирішення перерахованих проблем було реалізовано кілька спеціалізованих драйверів і мережевих стеків, до яких відносяться, наприклад, комерційне рішення Sniffer10G від Emulex і Myricom, а також відкрита розробка PF_RING компанії Ntop. Ці рішення використовують схему з кільцевих буфером, як більш ефективну, а також оптимізовані для багатопроцесорних і багатоядерних комп'ютерів.

Угруповання пакетів в потоки - досить стандартна і проста операція. Основна відмінність різних реалізацій даного функціоналу пов'язано з тим, які саме поля адресної інформації і як використовувати для ідентифікації потоку. Найбільш вживане визначення потоку було дано раніше. Так як воно використовує 5-ку полів як ключову інформацію для визначення належності конкретного пакета до конкретного потоку, то для його позначення і зазвичай використовують термін 5-tuple. Модуль, який відповідає за угруповання пакета зазвичай називають генератором потоків (flow generator). В процесі роботи даний модуль зберігає в пам'яті відображення відповідної ключової інформації на дані конкретних потоків.

Тема класифікації мережевого трафіку сама по собі є дуже великою. Перш ніж переходити до методів, якими вона здійснюється, перерахуємо варіанти класифікації за її результатами, тобто об'єктами, які виходять на виході даного алгоритму, їх властивостями і можливостям їх подальшої обробки. За цим критерієм, можна виділити три основні варіанти класифікації.

Тип трафіку не є достатньо змістовним способом класифікації і, як правило, або не підлягає подальшого аналізу, або піддається досить простій додатковій уточнюючій класифікації. Залежно від сфери застосування, типи можуть бути різними.

Використовується протокол прикладного рівня (protocol identification) є досить змістовним і може, використовуватися безпосередньо - наприклад, в системах збору статистики і моніторингу для підвищення рівня точності. Основним способом подальшої обробки є розбір протоколу, що включає дві основних функції - складання сесії прикладного рівня, в разі необхідності вилучення даних протоколу з окремих його полів (метаінформація рівня протоколу).

Додаток, що передає дані (application identification), дає максимально деталізований рівень класифікації. На цьому рівні можуть здійснюватися ті ж види обробки, що і на рівні протоколу прикладного рівня, а також вилучатись і інтерпретуватися дані (метаінформація) конкретного додатку, що відповідає більш високому рівню їх уявлення. Наприклад, поле типу «рядок», визначене на рівні протоколу, може відповідати «імені користувача» на рівні додатку.

З наведеного огляду можна зробити ряд висновків. Можна констатувати як кількісне, так і якісне зростання потреб в засобах аналізу трафіку. При цьому, незважаючи на величезне різноманіття конкретних рішень, що реалізують різні види аналізу, в основі більшості цих рішень лежить приблизно однакова схема.