УДК 681.3

АНАЛИЗ ТЕХНОЛОГИИ DEEP PACKET INSPECTION

© 2018 В. В. Воронин

Министерство иностранных дел Российской Федерации (г. Москва, Россия)

В статье дан анализ технологии Deep Packet Inspection. Отмечены два основных метода классификации трафика в компьютерных сетях. Показаны возможные сценарии работы системы.

Ключевые слова: компьютерная сеть, трафик, компьютерная программа, маршрутизация.

DPI – (Deep Packet Inspection) определяет технологию, на основе которой статистические данные накапливаются, сетевые пакеты проверяются и фильтруются относительно их содержимого [2-4]. Deep Packet Inspection имеет отличие от брандмауэров, поскольку анализируются не только заголовки в пакетах, но и содержимое трафика полным образом для уровней модели OSI от второго вверх. На базе Deep Packet Inspection обнаруживаются и блокируются вирусы, информация фильтровать, которая не соотносится с заданными критериями [6, 7].[1]

В рамках Deep Packet Inspection могут приниматься решения не только относительно содержимого пакетов, но и в случае косвенных признаков, которые характерны для соответствующих сетевых протоколов и программ. С этой целью применяют статистический анализ (он может быть связан с частотным анализом встреч соответствующих символов, длин пакетов и др.).

Провайдерами Deep Packet Inspection применяется с целью контролировать трафик, а в ряде случаев, для блокировки каких-то протоколов, например, BitTorrent. На базе Deep Packet Inspection есть возможности определения того, каким приложением были генерированы или получены данные, и, исходя из этого, выдача соответствующих действий. Еще, Deep Packet Inspection позволяет проводить процесс сбора подробных статистических данных по соединениям каждого пользователя.

Сетевой трафик неоднороден, он состоит из множества приложений, протоколов и сервисов [6, 11, 14]. Многие из этих приложений уникальны по требованиям к характеристикам сети, таким как скорость, задержки, джиттер. Удовлетворение этих требований – обязательное условие для того, чтобы приложение работало быстро и стабильно, а пользователи были удовлетворены качеством. Если в локальных сетях (LAN) с их высокой пропускной способностью проблем не бывает, то ограниченная ширина канала доступа в Интернет (WAN) требует тонкой настройки [8, 12, 13].

Классификация трафика – первый шаг, который помогает идентифицировать различные приложения и протоколы, передаваемые по сети. Вторым шагом является управление этим трафиком, его оптимизация и приоритизация. После классификации все пакеты становятся отмеченными по принадлежности к определенному протоколу или приложению, что позволяет сетевым устройствам применять политики обслуживания (QoS), опираясь на эти метки и флаги.

Основные понятия: классификация – идентификация приложений [1, 16] или протоколов; маркировка – процесс разметки пакетов для применения политик обслуживания на оборудовании.

Существуют два основных метода классификации трафика:

Классификация на основе блоков данных (Payload-Based Classification). Основывается на полях с блоками данных, таких как порты (Layer 4) OSI (отправитель и получатель или оба). Данный метод является наиболее распространенным, но не работает с зашифрованным и туннелированным трафиком. Классификация на основе статистического метода. Основывается на анализе поведения трафика (время между пакетами, время сеанса и т. п.). Универсальный подход к классификации трафика основывается на информации в заголовке IP-пакета – как правило, это IP-адрес (Layer 3), MAC-адрес (Layer 2), используемый протокол. Этот подход имеет ограниченные возможности, поскольку информация берется только из IP-заголовка, так же, как ограничены методы

Воронин Василий Владимирович – Министерство иностранных дел РФ, старший специалист, vass__7voorn4@yandex.ru.

Layer 4 – ведь далеко не все приложения используют стандартные порты. Более совершенную классификацию позволяет осуществить глубокий анализ пакетов (DPI). Это метод наиболее точный и надежный, его рассмотрим подробнее.

Технология Deep Packet Inspection.

Системы глубокого анализа трафика позволяют классифицировать те приложения и протоколы, которые невозможно определить на Layer 3 и Layer 4, например URL внутри пакета, содержимое сообщений мессенджеров, голосовой трафик Skype, p2p-пакеты BitTorrent.

Основным механизмом идентификации приложений в DPI является анализ сигнатур (Signature Analysis). Каждое приложение имеет свои уникальные характеристики, которые занесены в базу данных сигнатур. Сопоставление образца из базы с анализируемым трафиком позволяет точно определить приложение или протокол. Но так как периодически появляются новые приложения, то базу данных сигнатур [10, 15] также необходимо обновлять для обеспечения высокой точности идентификации.

Существуют несколько методов сигнатурного анализа:

- Анализ образца (Pattern analysis).
- Числовой анализ (Numerical analysis).
- Поведенческий анализ (Behavioral analysis).
- Эвристический анализ (Heuristic analysis).
- Анализ протокола/состояния (Protocol/state analysis).

Анализ образца. Некоторые приложения содержат определенные образцы (байты/символы/строки) в блоке данных пакета, которые можно использовать для идентификации и классификации. Причем образцы могут находиться в любом месте блока данных, это никак не влияет на процесс идентификации. Но так как не каждый пакет содержит в себе образец приложения, этот метод работает не всегда.

Числовой анализ. Числовой анализ изучает количественные характеристики пакетов, такие как размер блока данных, время отклика пакета, интервал между пакетами. Например, старая версия Skype (до версии 2.0) хорошо поддавалась такому анализу, потому что запрос от клиента имел размер 18 байт, а ответ, который он получал, – 11 байт. Поскольку анализ может быть распространен по пакетам сети магазинов, решение классификации могло бы занять больше времени. Одновременный анализ нескольких пакетов требует довольно много времени, что делает этот способ не самым эффективным.

Поведенческий и эвристический анализ.

Данный метод основывается на поведении трафика запущенного приложения. Пока приложение запущено, оно генерирует динамичный трафик, который также может быть идентифицирован и подвергнут маркировке. Например, BitTorrent генерирует трафик с определенной последовательностью пакетов, обладающих одинаковыми признаками (входящий и исходящий порт, размер пакета, число открываемых сессий в единицу времени), по поведенческой (эвристической) модели его можно классифицировать.

Поведенческий и эвристический анализ обычно применяют совместно, такие методы используют многие антивирусные программы для идентификации вирусов и червей.

Анализ протокола/состояния.

Протоколы некоторых приложений – это последовательность определенных действий. Анализ таких последовательностей позволяет достаточно точно идентифицировать приложение. Например, на запрос GET от FTP клиента обязательно следует соответствующий ответ сервера. Все больше приложений в Интернете начинают использовать механизмы шифрования трафика, что создает большие проблемы для любого из методов классификации. Система DPI не может заглянуть внутрь зашифрованного пакета для анализа содержимого, поэтому основными методами идентификации такого трафика являются поведенческий и эвристический анализ, но даже они могут определить далеко не все приложения. Новейший механизм, использующий оба эти метода одновременно, называется кластерным, и только он позволяет идентифицировать зашифрованный трафик. Так как ни один из описанных методов по отдельности не обеспечивает 100 %-ную классификацию трафика, лучшей практикой является использование их всех одновременно. Классификация трафика с дальнейшим применением политик качества обслуживания составляет одну из самых важных задач любого оператора связи. Использование современных систем DPI позволяет выполнять эту задачу с максимальной эффективностью и производительностью.

Схемы подключения DPI

Основных схем подключения устройства глубокого анализа трафика к оборудова-

нию оператора здесь две – это так называемая установка «в разрыв» (активная схема) и зеркалирование трафика (пассивная схема).

Схема установки «в разрыв». Этот вид подключения используется для реализации функционала любой системы DPI. В этом случае система анализа трафика подключается после граничного маршрутизатора в разрыв uplink. Преимуществом такой схемы является то, что через DPI проходит абсолютно весь трафик. Это позволяет осуществлять приоритизацию, а также настраивать уведомления, кеширование и другие функции. Однако такой тип подключения обладает существенным недостатком: устройство DPI становится точкой отказа – если оно выходит из строя, связь полностью обрывается. Есть способы решения этой проблемы:

1. Использовать в составе системы DPI bypass-устройство, которое в случае выхода из строя основного фронтенда начнет «гнать» трафик через себя (анализ трафика проводиться не будет).

2. Использовать резервную платформу DPI, которая бы осуществляла фильтрацию трафика в случае выхода из строя основной.

Схема зеркалирования трафика. Зеркалирование трафика осуществляется через SPAN-порты или оптические сплиттеры. При такой схеме возможен анализ истории посещений в реальном времени, переадресация запросов блокировки, кеширование и работа с бонусными программами. Плюсами этого варианта подключения являются минимальные изменения в структуре действующей сети [9] и отсутствие необходимости использовать bypass-карту. В этом случае есть возможность снимать аналитику с трафика, подключить кеш-сервер и «зеркалить» трафик на оборудование СОРМ, но весь функционал системы DPI реализовать не получится.

Сценарии использования системы.

1. Анализ и классификация трафика.

Ведение статистики позволяет оператору связи классифицировать проходящий через него трафик и знать какой тип трафика в какое время преобладает. Дает возможность составлять отчет по выбранным дням в удобных схемах и графиках.

2. Приоритизация трафика.

Дает возможность оператору связи грамотно осуществлять политику QoS для определенных групп протоколов. Тем самым улучшая качество обслуживания критичных к пингу приложений.

3. Оптимизация аплинков.

Уменьшение пропускной способности «прожорливых протоколов» во время кратковременного роста трафика.

4. Распределение канала между абонентами.

Гибкое ограничение пропускной способности определенных групп трафика (ограничивать Torrent или, наоборот, давать больше скорости на видео). На основании пользовательских предпочтений.

5. Кэширование.

Использование кеш-сервера в связке с системой DPI. Предполагает отдачу абоненту трафика из кеша сервера снижая загруженность своей сети и увеличивая скорость загрузки контента у абонента.

6. Поведенческая оценка абонентов.

Сбор статистической информации о уникальных предпочтениях пользователя (время использования, группа сайтов, группа протоколов). С целью улучшения качества обслуживания, увеличения числа абонентов и дохода компании.

7. Уведомление абонентов.

Функция, которая позволяет оператору передавать сообщения абоненту во время работы в Интернете. Пользователь вводит адрес сайта, который хочет посетить, и видит в браузере сообщение от оператора, сменяемое через несколько секунд запрашиваемой страницей.

8. Запрещение ресурсов (белый и черный списки).

Происходит блокировка интернет контента по требованию уполномоченных структур. А также при исчерпании средств на счету абонента переадресация на страницу оплаты разрешив доступ на страницу платежных систем и ограничив ко всем остальным сайтам.

9. Защита, перехват трафика, предфильтр СОРМ.

Так как DPI пропускает через себя и фильтрует весь трафик, защита абонентов и вычислительных систем в облаке становится для нее одной из непосредственных задач. Основными направлениями защиты являются:

1) Спам-боты (выявляются на основе анализа SMTP трафика).

2) DoS- и DDoS-атаки (выявляются по аномалиям трафика).

3) Заражение червями (выявляется по сигнатурам).

Вывод. Рассмотренная технология является достаточно перспективной и ее следует иметь ввиду при разработке современных приложений.

ЛИТЕРАТУРА

1. Балашова И. Ю. Разработка интегрированных средств представления знаний в интеллектуальной системе поддержки жизненного цикла программных продуктов / И. Ю. Балашова, Е. А. Дзюба, Е. Н. Прошкина // Моделирование, оптимизация и информационные технологии. – 2018. – Т. 6. – № 1 (20). – С. 357-367.

2. Варианты использования DPI. Часть 1 (Электронный ресурс – https://vasexperts.ru/blog/varianty-ispolzovaniya-dpi-chast-1/).

3. Варианты использования DPI. Часть 2 (Электронный ресурс -https://vasexperts.ru /blog/varianty-ispolzovaniya-dpi-chast-2/).

4. Введение в DPI: Состав системы и схемы подключения (Электронный ресурс - https://habr.com/company/vasexperts/blog/313554/).

5. Дубровин М. Г. Модели и методы проактивного мониторинга ИТ-СИСТЕМ / М. Г. Дубровин, И. Н. Глухих // Моделирование, оптимизация и информационные технологии. – 2018. – Т. 6. – № 1 (20). – С. 314-324.

6. Deep packet inspection (Электронный ресурс – https://ru.wikipedia.org/wiki/ Deep_packet_inspection)

7. Классификация трафика и Deep Packet Inspection (Электронный ресурс – https://vasexperts.ru/blog/klassifikatsiya-trafika-i-deep-packet-inspection/)

8. Львович Я. Е. Исследование методов оптимизации при проектировании систем радиосвязи / Я. Е. Львович, И. Я. Львович, А. П. Преображенский, С. О. Головинов // Теория и техника радиосвязи. – 2011. – № 1. – С. 5-9.

9. Lvovich I. Ya. The analysis of scattering electromagnetic waves with use of paraller computing / I. Ya. Lvovich, A. P. Preobrazhenskiy, O. N. Choporov, K. V. Kaydakova // International Siberian Conference on Control and Communications, SIBCON 2015 – Proceedings 2015. – С. 7147133.

10. Недосекин Д. А. Многовариантный выбор при управлении развивающимися системами / Д. А. Недосекин // Моделирование, оптимизация и информационные технологии. – 2018. – Т. 6. – № 1 (20). – С. 346-356.

11. Поначугин А. В. Моделирование системы радиодоступа в мультисервисных сетях связи / А. В. Поначугин, И. В. Гусев // Моделирование, оптимизация и информационные технологии. – 2018. – Т. 6. – № 1 (20). – С. 118-130.

12. Попов А. А., Кузьмина А. О. Формирование набора компонентов программного обеспечения для выполнения обязанностей диспетчера аварийно-диспетчерской службы жилищно-коммунального хозяйства // Моделирование, оптимизация и информационные технологии. – 2018. – Т. 6. – № 1 (20). – С. 153-175.

13. Преображенский А. П., Характеристики распространения радиоволн в подземных беспроводных системах связи / А. П. Преображенский, А. А. Хромых // Моделирование, оптимизация и информационные технологии. – 2013. – № 2 (2). – С. 5.

14. Преображенский А. П. О применении расчетно-экспериментального подхода при исследовании распространения волн Wi-Fi внутри помещения / А. П. Преображенский // Вестник Воронежского института высоких технологий. – 2014. – № 12. – С. 71-72.

15. Попова Н. А. Решение задачи распознавания лиц с использованием алгоритмов машинного обучения / Н. А. Попова, М. А. Назаров, М. В. Власов // Моделирование, оптимизация и информационные технологии. – 2018. – Т. 6. – № 1 (20). – С. 408-415.

16. Тишуков Б. Н. Повышение эффективности функционирования объектов со структурновариативной формой управления на основе оптимизационного моделирования / Б. Н. Тишуков // Моделирование, оптимизация и информационные технологии. – 2018. – Т. 6. – № 1 (20). – С. 288-298.

ANALYSIS OF THE TECHNOLOGY DEEP PACKET INSPECTION

© 2018 V. V. Voronin

Ministry of foreign Affairs (Moscow, Russia)

The analysis of the technology of Deep Packet Inspection is carried out. There are two main methods of classification of traffic in computer networks. Possible scenarios of the system operation are shown.

Key words: computer network, traffic, computer program, routing.