РАЗРАБОТКА ПРОГРАММНОГО АНАЛИЗАТОРА СЕТЕВОГО ТРАФИКА НА ОСНОВЕ КОЛИЧЕСТВЕННЫХ ХАРАКТЕРИСТИК ПОТОКА ТРАНСПОРТНОГО УРОВНЯ

М.Н. Беленькая, Д.О. Прохоров, Н.В. Трофлянина, С.А. Фомин

В современных условиях, когда по сетям передачи данных передаются огромные объёмы информации, остро необходимы программные продукты, позволяющие быстро анализировать сетевой трафик с целью определения его природы. В данной работе представлен способ определения природы передаваемого трафика на основе количественных характеристик потока транспортного уровня. Отличительная черта разработанного способа – его легковесность и нетребовательность к ресурсам машины, на которой выполняется анализ. Разработано приложение, выполняющее поставленную задачу в реальном времени.

Ключевые слова: анализ трафика, потоки трафика, машинное обучение.

DEVELOPMENT OF SOFTWARE ANALYZER NETWORK TRAFFIC BASED ON QUANTITATIVE CHARACTERISTICS OF TRANSPORT STREAM

Belenkaya M., Prokhorov D., Troflyanina N., Fomin S.

In modern conditions, when data networks to transfer huge amounts of information, badly needed software to quickly analyze network traffic to determine its nature. This paper presents a method of determining the nature of the traffic on the basis of quantitative transport layer flow characteristics. A distinctive feature of the developed method is its lightness and undemanding to resources of the machine on which the analysis is performed. It developed an application that performs the task in real time.

Keywords: traffic analysis, traffic flows, machine learning.

Введение

С каждым годом всё больше растут объёмы сетевого трафика, передаваемого по сети Интернет. Для решения многих прикладных задач требуется проводить анализ этого трафика – например, для организации классов обслуживания, для обнаружения нежелательного и запрещённого трафика и т. д.

Одним из средств, позволяющих проводить анализ передаваемых данных, является глубокий анализ пакетов (DPI). К достоинствам этого способа можно отнести высокую точность анализа, к недостаткам – низкую скорость, что критично при современных объёмах передаваемого трафика.

Другой распространённый способ – анализ TCP- и UDP-портов источника и получателя и определение приложения на основе списка широко известных портов. Недостаток способа в его недостоверности (известный порт может быть использован другим приложением) и в том, что по одному и тому же порту может передаваться трафик разной природы.

В данной работе предлагается быстрый способ анализа трафика, основанный на количественных характеристиках потока транспортного уровня. Этот способ намного быстрее, чем DPI, поскольку при его осуществлении не требуется смотреть и записывать данные прикладного уровня, достаточно запоминать лишь некоторые характеристики потока. Также предложенный способ не использует априорных знаний об использовании приложениями известных номеров TCP- и UDP-портов.

Целью работы ставится разработка способа разделения сетевого трафика на классы обслуживания с целью повышения качества обслуживания интернет-провайдера, а также разработка программного обеспечения, способного выполнять поставленную задачу в реальном времени.

1. Классы обслуживания

При разработке классов обслуживания за основу было взято разделение, предложенное телекоммуникационным стандартизирующим сектором Международного Союза Электросвязи (ITU-T), имеющее название Y.1541. Это разделение содержит 6 классов трафика, каждый из которых в разной степени требователен к одной из четырёх характеристик:

1) Задержка (верхняя граница для среднего значения задержки).
2) Джиттер (верхняя граница задержки для квантиля 99,9 %).
3) Потери (верхняя граница на вероятность потери пакета).
4) Ошибки (верхняя граница на вероятность ошибки в пакете).

Так как требования к потерям и ошибкам одинаковы почти у всех классов (они отсутствуют лишь у шестого, «неопределённого» класса»), на основе предложенных *ITU* характеристик были выделены следующие 3 типа сетевых приложений:

- Тип 0: не чувствительные к задержкам и джиттеру. Примеры: браузеры при просмотре веб-страниц (без мультимедиа), BitTorrent.
- Тип 1: чувствителен только к задержкам, но не к джиттеру. Пример: браузеры при просмотре потокового мультимедиа.
- Тип 2: чувствительные и к задержкам, и к джиттеру. Примеры: Skype, Viber при осуществлении аудио- и видеосвязи реального времени.

В данной классификации оказывается, что трафик, генерируемый одним и тем же приложением ОС, может относиться к разным классам обслуживания. В этой связи необходимо уточнить понятие «приложение». В рамках данной работы под приложением понимается владелец одного открытого сокета. Например, приложение ОС chrome.exe может одновременно держать открытым сокет на некотором TCP-порту и принимать через этот порт гипертекстовую разметку веб-страницы (HTML), а также держать открытым сокет на некотором UDP-порту и принимать через этот сокет видеоданные (с такого сайта, как YouTube). В этой ситуации будем считать, что работают два отдельных приложения: одно типа 0 (гипертекст), другое типа 1 (мультимедиа).

2. Обучающая выборка

Для проведения анализа отобраны следующие виды трафика:

- Тип 0: DNS, BitTorrent, HTML поверх HTTP(S).
- Тип 1: видео и аудио поверх HTTP(S) и QUIC.
- Тип 2: видео и аудио в ПО Skype.

Были выбраны именно эти виды трафика по той причине, что именно такой трафик генерируют большинство пользователей сети Интернет.

Было собрано более 3 Гб трафика приведённых выше видов. Для выделения этих видов среди прочего трафика использована библиотека nDPI, осуществляющая глубокий анализ пакетов. Однако nDPI определяет лишь протокол прикладного уровня, а этого оказывается недостаточно для разделения трафика на необходимые классы, так что кроме DPI был проведён дополнительный анализ трафика каждого протокола. Итоговое распределение потоков по классам представлено в таблице 1.

ТАБЛИЦА 1. Критерии отбора трафика для обучающей выборки

Тип трафика	Тип приложения	Критерий разделения
DNS	0	–
BitTorrent	0	–
HTTP.text	0	Анализ HTTP-заголовка, разделение по MIME-типу
HTTP.image	0	
HTTP.audio	1	
HTTP.video	1	
SSL.multimedia	1	Эвристическое разделение: мультимедиа > 1 Мб полезной нагрузки
SSL.other	0	
QUIC.multimedia	1	
QUIC.other	0	
Skype.realtime	2	Эвристическое разделение: realtime > 50 пакетов в секунду и > 3 секунд длительности
Skype.other	0	

3. Обучение классификатора

Следующий шаг – выделение количественных характеристик потока и применение алгоритмов машинного обучения для создания классификатора.

При выделении количественных характеристик потока следует пояснить некоторые используемые понятия:

- порция данных – объём полезной нагрузки, переданный от одной стороны к другой, не прерываемый полезной нагрузкой с противоположной стороны;
- инициатор – инициатор соединения;
- адресат – адресат соединения.

Полный список выделенных характеристик представлен в таблице 2.

ТАБЛИЦА 2. Полный список выделяемых характеристик потока

Характеристика	**Пояснение**
Средний размер сегмента со стороны инициатора	Среднее значение – статистическая оценка математического ожидания
Средний размер сегмента со стороны адресата	
Средний размер порции данных со стороны адресата	
Средний размер порции данных со стороны инициатора	
СКО размера порции данных со стороны инициатора	Среднеквадратическое отклонение – статистическая оценка дисперсии
СКО размера порции данных со стороны адресата	
СКО размера сегмента со стороны инициатора	
СКО размера сегмента со стороны адресата	
Среднее число сегментов на порцию данных со стороны инициатора	Общее число пакетов, делённое на общее число порций данных
Среднее число сегментов на порцию данных со стороны адресата	
КПД инициатора	Общее количество полезной нагрузки, делённое на общее количество байт
КПД адресата	
Соотношение переданных байт	Во сколько раз адресат передаёт больше, чем инициатор
Соотношение переданной полезной нагрузки	
Соотношение переданных сегментов	
Размеры первых 4 порций данных	2 со стороны инициатора, 2 со стороны адресата
Размеры первых 4 сегментов данных	Не учитывая TCP handshake

Значения характеристик определялись по первой 1 000 пакетов каждого потока.

Было испробовано несколько алгоритмов машинного обучения, однако лучший результат по Ф-мере показал алгоритм «Случайный Лес». Для обучения классификатора было использовано две трети выборки, для проверки его производительности – оставшаяся треть. Результаты работы алгоритма представлены в таблице 3.

ТАБЛИЦА 3. Проверка классификатора на тестовой выборке

	Предсказанные классы			
Реальные классы		0	1	2
	0	1624	1	0
	1	0	36	0
	2	0	0	3

В приведённой таблице значение каждой ячейки – это количество случаев, когда классификатор причислил поток к классу, указанному в заголовке столбца, а на самом деле трафик принадлежат классу, указанному в заголовке строки. Таким образом, главная диагональ – случаи верной классификации, всё остальное – разного рода ошибки.

Кроме того, была предпринята попытка обучить классификатор определять не только тип приложения, но и природу самого трафика. Результаты представлены в таблице 4.

Также было разработано программное обеспечение, способное в реальном времени разделять трафик на потоки, вести учёт количественных характеристик каждого потока и на основе собранных характеристик делать выводы о природе передаваемого трафика. На рисунке представлен скриншот разработанного приложения.

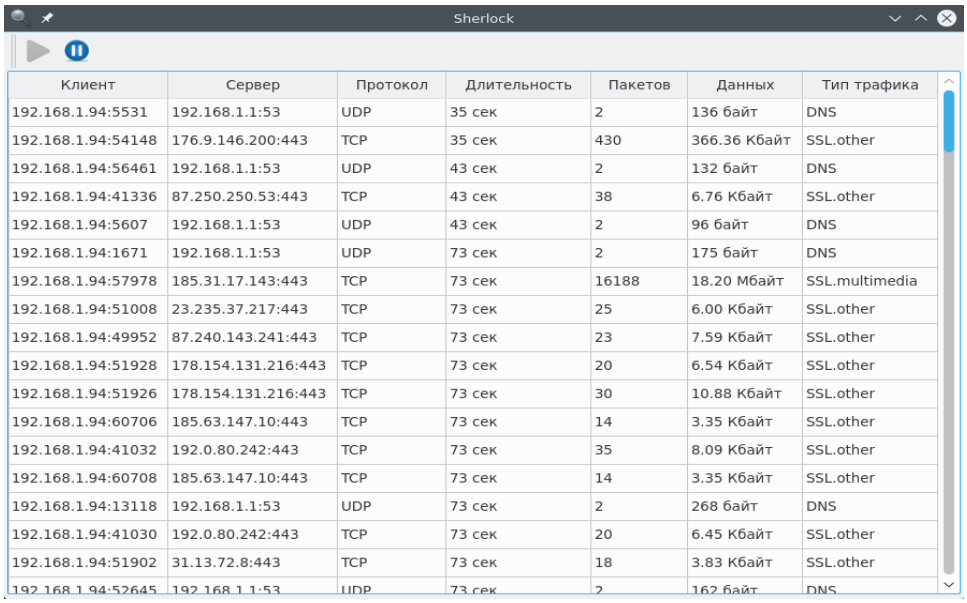
ТАБЛИЦА 4. Результат расширенного обучения классификатора

	Предсказанные классы											
Усл. № класса	1	2	3	4	5	6	7	8	9	10	11	12
1. BitTorrent	20											
2. DNS		1019										
3. HTTP.audio			6			2						
4. HTTP.image				139	1							
5. HTTP.text				4	74					3		
6. HTTP.video						2						
7. QUIC.multimedia							1					
8. QUIC.other								38				
9. SSL.multimedia			3			1			23	1		
10. SSL.other				1						239		
11. Skype.other											88	
12. Skype.realtime												3

Выводы

В результате эксперимента была доказана принципиальная возможность определения класса сетевого приложения на основе количественных характеристик потока транспортного уровня. Более того, было показано, что с некоторыми погрешностями возможно даже определять протокол 7-го уровня и природу передаваемых данных без анализа содержимого пакетов. Был разработан программный продукт, способный выполнять анализ трафика на основе количественных характеристик потока в реальном времени. Дальнейшее развитие предложенных идей возможно в нескольких направлениях. Во-первых, следует

усовершенствовать методы отделения целевого трафика, чтобы устранить эвристические критерии. Во-вторых, представленные идеи требуется проверить на большем объёме трафика, ведь собранные 3 Гб не могут претендовать на полную репрезентативность.



Клиент	Сервер	Протокол	Длительность	Пакетов	Данных	Тип трафика
192.168.1.94:5531	192.168.1.1:53	UDP	35 сек	2	136 байт	DNS
192.168.1.94:54148	176.9.146.200:443	TCP	35 сек	430	366.36 Кбайт	SSL.other
192.168.1.94:56461	192.168.1.1:53	UDP	43 сек	2	132 байт	DNS
192.168.1.94:41336	87.250.250.53:443	TCP	43 сек	38	6.76 Кбайт	SSL.other
192.168.1.94:5607	192.168.1.1:53	UDP	43 сек	2	96 байт	DNS
192.168.1.94:1671	192.168.1.1:53	UDP	73 сек	2	175 байт	DNS
192.168.1.94:57978	185.31.17.143:443	TCP	73 сек	16188	18.20 Мбайт	SSL.multimedia
192.168.1.94:51008	23.235.37.217:443	TCP	73 сек	25	6.00 Кбайт	SSL.other
192.168.1.94:49952	87.240.143.241:443	TCP	73 сек	23	7.59 Кбайт	SSL.other
192.168.1.94:51928	178.154.131.216:443	TCP	73 сек	20	6.54 Кбайт	SSL.other
192.168.1.94:51926	178.154.131.216:443	TCP	73 сек	30	10.88 Кбайт	SSL.other
192.168.1.94:60706	185.63.147.10:443	TCP	73 сек	14	3.35 Кбайт	SSL.other
192.168.1.94:41032	192.0.80.242:443	TCP	73 сек	35	8.09 Кбайт	SSL.other
192.168.1.94:60708	185.63.147.10:443	TCP	73 сек	14	3.35 Кбайт	SSL.other
192.168.1.94:13118	192.168.1.1:53	UDP	73 сек	2	268 байт	DNS
192.168.1.94:41030	192.0.80.242:443	TCP	73 сек	20	6.45 Кбайт	SSL.other
192.168.1.94:51902	31.13.72.8:443	TCP	73 сек	18	3.83 Кбайт	SSL.other
192.168.1.94:52645	192.168.1.1:53	UDP	73 сек	2	162 байт	DNS

Рисунок. Скриншот разработанного приложения

Список используемых источников

1. Lim Y., Internet Traffic Classification Demystified: On the Sources of the Discriminative Power, конференция CoNEXT 2010.

ИСТОРИЯ РАЗВИТИЯ ОТЕЧЕСТВЕННОЙ СПУТНИКОВОЙ СВЯЗИ

В.М. Березянская, А.Ю. Мартынов, С.И. Савинский, А.К. Сагдеев

Статья посвящена истории развития отечественной спутниковой связи, идею которой впервые предложил Артур Кларк в 1945 г. В статье рассмотрен первый искусственный спутник земли, его запуск и усовершенствование, создание различных систем спутниковой связи, а также полет первого человека в космос.

Ключевые слова: Артур Кларк, спутник, спутниковая связь, радиоретранслятор.

THE HISTORY OF DOMESTIC SATELLITE COMMUNICATION DEVELOPMENT

Berezyanskaya V., Martynov A., Savinskiy S., Sagdeev A.

The article describes the history of development of the domestic satellite communication, where the idea was first proposed by Arthur C. Clarke in 1945. The first artificial satellite to be