УДК 004.7

ИССЛЕДОВАНИЕ МЕТОДОВ СБОРА СТАТИСТИЧЕСКИХ ДАННЫХ О ТРАФИКЕ В IP-СЕТЯХ ПЕРЕДАЧИ ДАННЫХ

Михеев А.В.

ФГАОУ ВПО Уральский федеральный университет имени первого Президента России Б.Н. Ельцина, Екатеринбург, Россия (620000 Россия, г.Екатеринбург, ул. Мира, 32), e-mail: a.v.mikheev@hotmail.com

Аннотация: Настоящая статья посвящена исследованию нескольких различных методов сбора статистических данных, применительно к IP-сетям начиная с самых примитивных и заканчивая наиболее фукциональными – коммандная строка, SNMP, прокси-сервер, NetFlow, DPI. Все эти методы так или иначе применяются во всех современных сетях передачи данных. Были рассмотрены различия между методами и применимость каждого из них на практике.

Ключевые слова: IP-трафик, NetFlow, коммандная строка, SNMP, прокси-сервер, DPI.

RESEARCH METHODS OF COLLECTING TRAFFIC STATISTICS IN IP DATA NETWORKS

Mikheev A.V.

Federal Autonomous Educational Institution of Higher Professional Education Ural Federal University named after the first President of Russia B. N. Yeltsin, Yekaterinburg, Russia (620000 Russia, Yekaterinburg, Mira street, 32), e-mail mikheev@hotmail.com

This article is devoted to the study of several different statistical techniques, applied to IP-based networks – from the most primitive and ending with the most functionality - the command line, SNMP, proxy server, NetFlow, DPI. All these methods are used anyway in all modern data networks. Have been considered the differences between the methods and applicability of each of them practically.

Key words: IP traffic, NetFlow, CLI, SNMP, proxy server, DPI.

Введение

Одной из актуальных проблем, с которыми сталкивается сетевой инженер, является учет и последующий анализ данных, передаваемых по сети, т.к. на сегодняшний день отсутствует какой-либо единый стандартный метод для учета сетевого трафика и наглядного выявления всех потоков данных и их основных характеристик.

Цель исследования

Рассмотреть пять разных методов сбора статистических данных о сетевом трафике, привести сравнение и анализ методов.

Основные характеристики IP трафика

Для того чтобы любая сеть передачи данных работала качественно и безопасно требуется вести статистику по передаваемому трафику и мониторить некоторые параметры для последующего анализа и оперативного управления сетью:

– IP адрес источника, IP адрес отправителя;
– TCP/UDP порт источника, TCP/UDP порт отправителя;
– задержка (мин/макс/сред) - промежуток времени, требуемый для передачи пакета через сеть;
– джиттер - задержка между двумя последовательными пакетами;
– процент потеряных пакетов;
– объем трафика, передаваемый в секунду (Mb/s);
– количество IP пакетов, передаваемое в секунду (pps);
– длина IP пакета (мин/макс/сред);
– выявление и анализ высокоуровневых прикладных протоколов.

Методы сбора статистических данных

Интерфейс командной строки сетевого оборудования

В самом простейшем случае сетевому администратору может потребоваться узнать *текущую нагрузку* на определенный сетевой интерфейс или канал связи, то есть нужно узнать ***объем трафика, передаваемый в секунду*** и ***количество IP пакетов, передаваемое в секунду***. Это можно сделать из командной строки без каких-либо дополнительных инструментов, лишь программно подключившись к оборудованию (рис. 1).

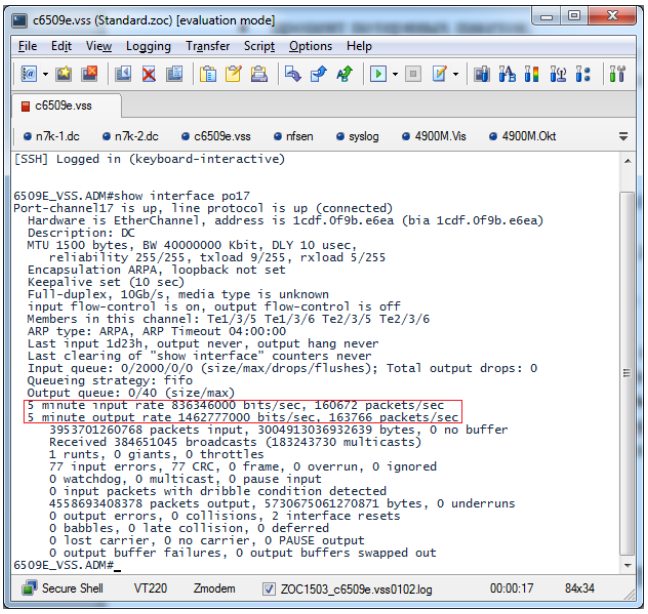


Рисунок 1. Консоль управления коммутатором

Такой метод подходит лишь для оперативного управления сетью и не подходит для ведения статистики по трафику, его анализа и прогнозирования каких-либо событий из-за недостатка необходимой информации (табл. 1).

Таблица 1.Характеристики, получаемые непосредственно с оборудования

IP адрес источника, IP адрес отправителя;	нет
TCP/UDP порт источника, TCP/UDP порт отправителя	нет
Задержка (мин/макс/сред)	нет
Джиттер	нет
Процент потерянных пакетов	нет
Объем трафика, передаваемый в секунду	да
Количество IP пакетов, передаваемое в секунду	да
Длина IP пакета (мин/макс/сред)	нет
Выявление прикладных протоколов	нет

Simple Network Management Protocol

SNMP – протокол, используемый для управления устройствами в IP-сетях. Протокол используется в системах сетевого управления для контроля, подключенных к сети устройств на предмет условий, которые требуют внимания администратора. Включает в себя схему баз данных и набор объектов данных, к которым можно обращаться с помощью специальных сетевых запросов. Другими словами SNMP предоставляет данные для управления в виде переменных, описывающих конфигурацию управляемой системы. И эти переменные могут быть запрошены (а иногда и заданы) по сети управляющими приложениями.

Одним из распространенных приложений, выполняющих эти задачи и наглядно представляющих информацию, является Cacti (рис. 2). Это достигается за счет периодического опроса оборудования по целому ряду параметров, но из параметров о трафике таким образом можно получить ***объем трафика, передаваемый в секунду*** и ***количество IP пакетов, передаваемое в секунду***.

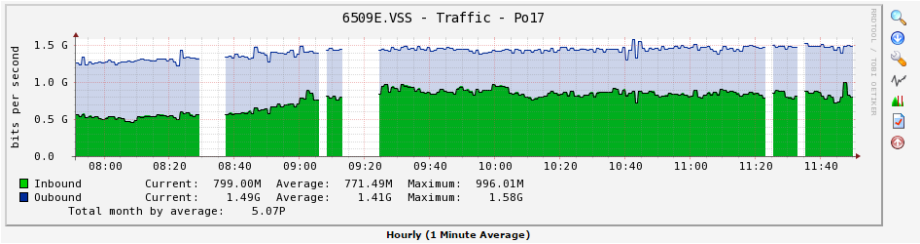


Рисунок 2. Интерфейс Cacti

Таблица 2.Характеристики, получаемые с помощью SNMP

IP адрес источника, IP адрес отправителя;	нет
TCP/UDP порт источника, TCP/UDP порт отправителя	нет
Задержка (мин/макс/сред)	нет
Джиттер	нет
Процент потерянных пакетов	нет
Объем трафика, передаваемый в секунду	да
Количество IP пакетов, передаваемое в секунду	да
Длина IP пакета (мин/макс/сред)	нет
Выявление прикладных протоколов	нет

Данный метод является пассивным (по отношению к конечным клиентам) и отлично подходит для мониторинга загрузки каналов связи и сетевых интерфейсов. Также

накопленная таким образом информация может быть использована при расследовании каких-либо инцидентов информационной безопасности, выявлении нештатных ситуаций из-за технических сбоев, при прогнозировании и анализе загрузки дорогостоящих (по деньгам или по времени) каналов связи.

Прокси-сервер

Отдельный сервер, позволяющий клиентам выполнять косвенные запросы к другим сетевым службам, то есть сначала клиент подключается к прокси-серверу и запрашивает какой-либо ресурс (например, e-mail), расположенный на другом сервере. Затем прокси-сервер либо подключается к указанному серверу и получает ресурс у него, либо возвращает ресурс из собственного кэша (в случаях, если прокси имеет свой кэш).

Основной недостаток – на всех компьютерах в сети требуется указать адрес прокси-сервера и установить клиентский агент (т.к. в стандартных операционных системах семейства Windows нет возможности перенаправить весь трафик на прокси-сервер.), что в крупных административно разделенных сетях бывает очень затруднительно или невозможно.

В качестве примера рассмотрим прокси-сервер Traffic Inspector, разрабатываемый российской компанией «Смарт-Софт».

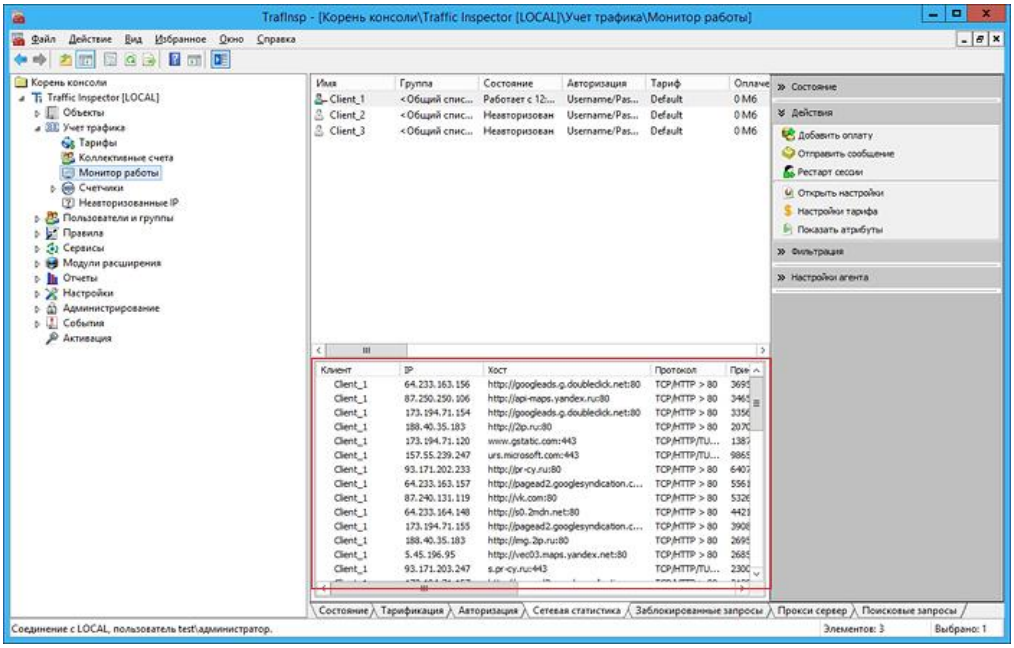


Рисунок 3. Анализ всего проксируемого трафика

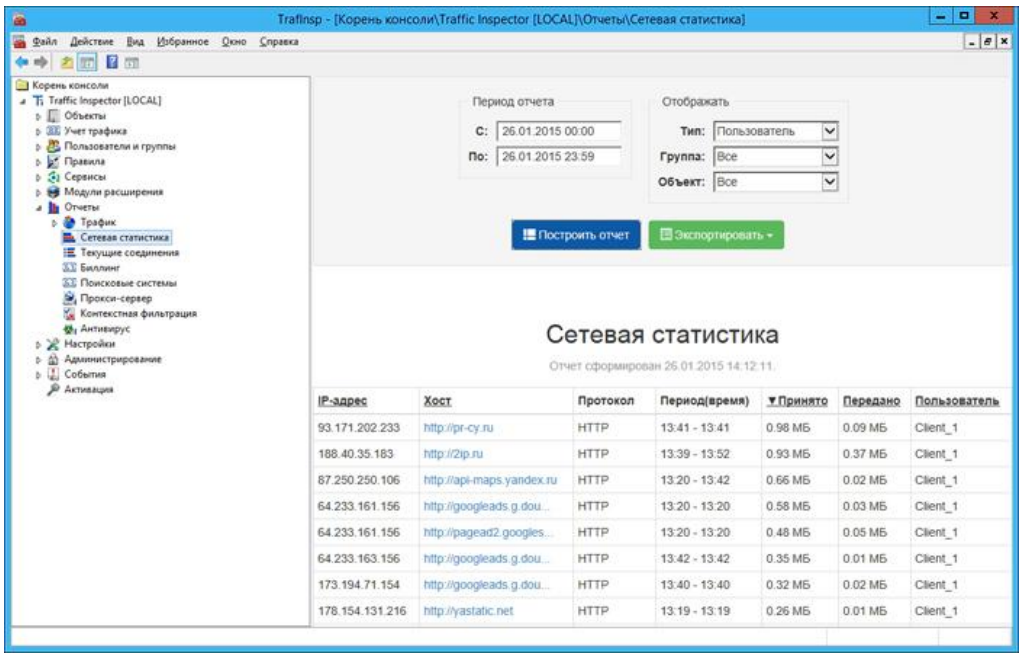


Рисунок 4. Встроенная возможность формирования отчетов

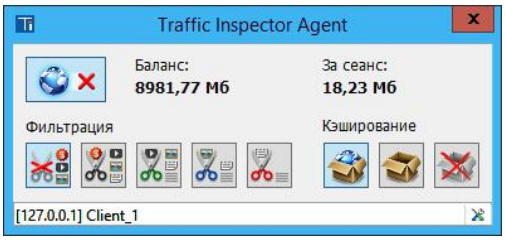


Рисунок 5. Клиентский агент

Таблица 3. Характеристики, получаемые с помощью прокси-сервера

IP адрес источника, IP адрес отправителя;	да
TCP/UDP порт источника, TCP/UDP порт отправителя	да
Задержка (мин/макс/сред)	нет
Джиттер	нет
Процент потерянных пакетов	нет
Объем трафика, передаваемый в секунду	да
Количество IP пакетов, передаваемое в секунду	нет
Длина IP пакета (мин/макс/сред)	нет
Выявление прикладных протоколов	да

Проксирование это действительно хороший метод для анализа трафика, если его недостатки не являются критичными (например, интернет провайдеры все чаще нуждаются в анализе проходящего трафика, но доступа к клиентским устройствам не имеют). Сама идея позволяет делать с трафиком практически все что угодно, но очень часто все упирается в реализацию и вычислительные возможности сервера, поэтому статистика ведется только по самым необходимым параметрам.

NetFlow

NetFlow — сетевой протокол, предназначенный для учёта сетевого трафика, разработанный компанией Cisco Systems. Поддерживается не только оборудованием Cisco, но и многими другими устройствами (в частности, Juniper и Enterasys). Также существуют свободные реализации для UNIX-подобных систем.

Существует несколько версий протокола, наиболее распространёнными из которых являются версии 5 и 9. На основе версии 9 также был разработан открытый стандарт под названием IPFIX (Internet Protocol Flow Information eXport, экспорт информации о потоках IP).

Используется распределенная архитектура и для сбора информации о трафике по протоколу NetFlow требуются следующие компоненты:

1. Сенсор

Собирает статистику по проходящему через него трафику. Обычно это L3-коммутатор или маршрутизатор, хотя можно использовать и отдельно стоящие сенсоры, получающие данные путем зеркалирования порта коммутатора.

2. Коллектор

Собирает получаемые от сенсора данные и помещает их в хранилище.

3. Анализатор

Анализирует собранные коллектором данные и формирует пригодные для чтения человеком отчёты (часто в виде графиков). NetFlow использует UDP для передачи данных о трафике коллектору. Как правило, коллектор «слушает» порт 2055, 9555 или 9995. Роли коллектора и анализатора могут быть объединены в одном устройстве/программе.

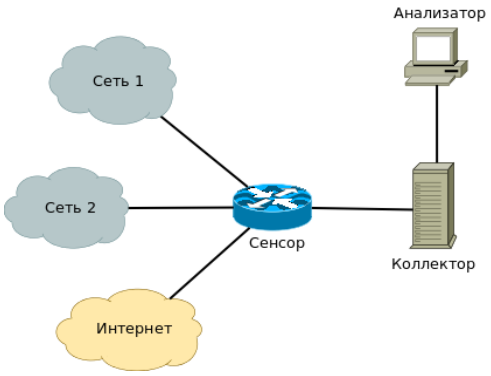


Рисунок 6. Архитектура NetFlow

Сенсор может выделять из проходящего трафика потоки. Когда сенсор определяет, что поток закончился (по изменению параметров пакетов, либо по сбросу TCP-сессии), он отправляет информацию в коллектор. В зависимости от настроек он также может периодически отправлять в коллектор информацию обо все еще идущих потоках.

Для экономии ресурсов процессора также применяется «sampled NetFlow». В этом случае сенсор анализирует не все, а каждый n-ый пакет, где n может быть заданным административно или выбираемым случайным образом. При использовании sampled NetFlow получаемые значения являются не точными, а оценочными. Обычно такой функционал применяется в центрах обработки информации, где большие объемы передаваемого трафика могут существенно усложнить задачу сбора статистики, но мониторинг необходим.

Рассмотрим NetFlow – в качестве сенсора возьмем набор устройств Cisco, в качестве коллектора и одновременно анализатора программу nfdump с веб-интерфейсом nfsen (рис.7, 8).

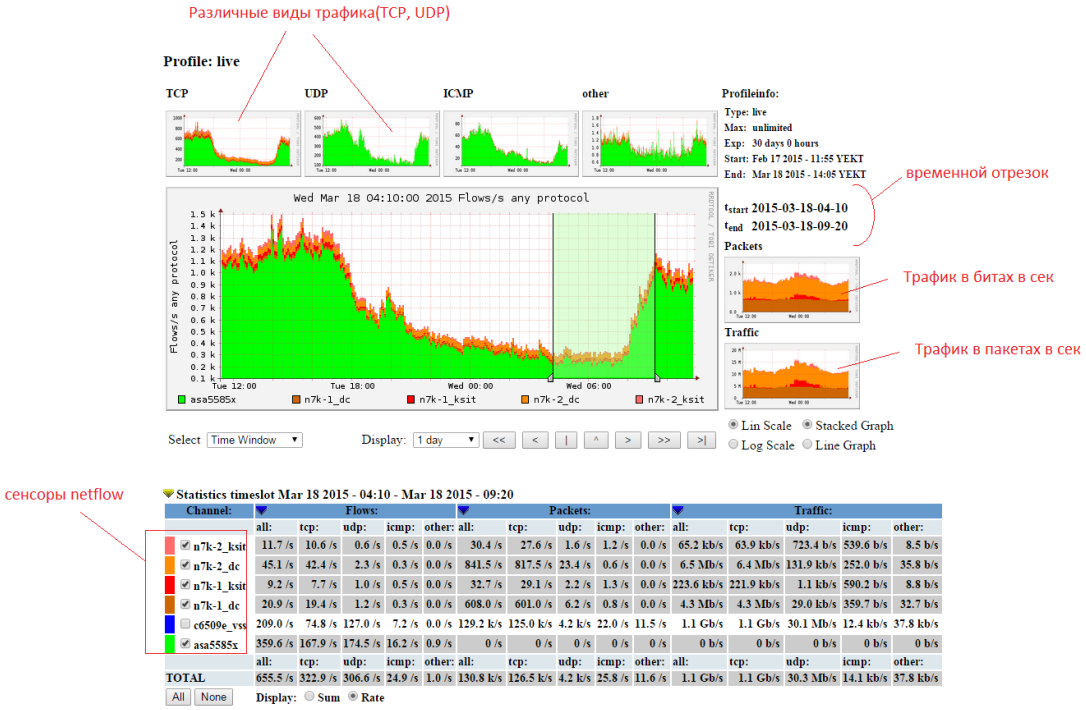
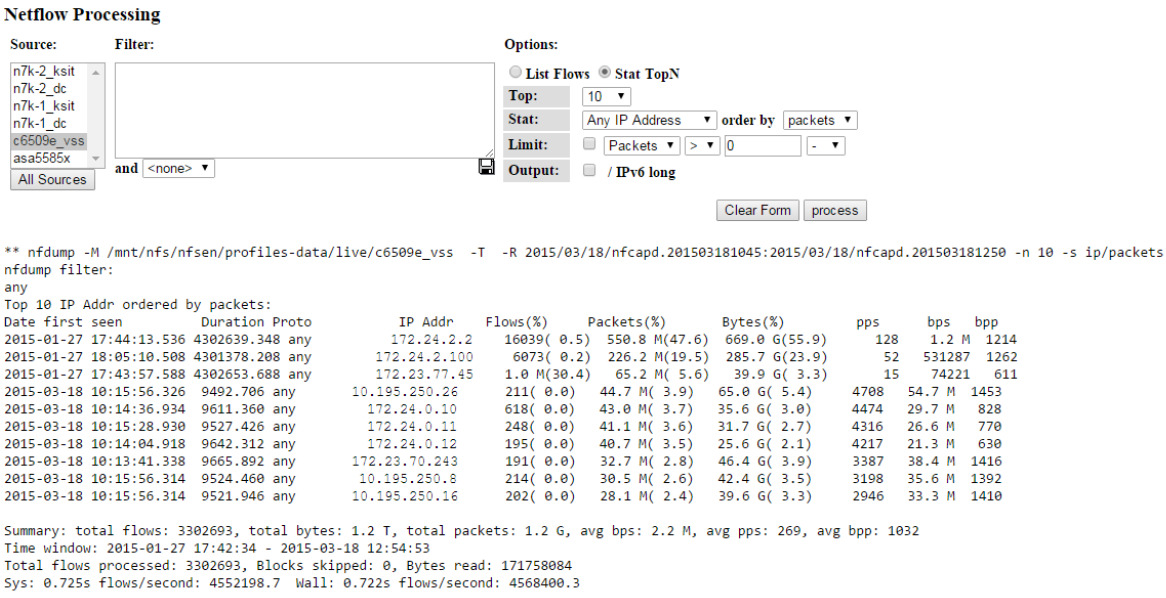


Рисунок 7. Общий интерфейс nfsen




```
** nfdump -M /mnt/nfs/nfsen/profiles-data/live/c6509e_vss  -T  -R 2015/03/18/nfcapd.201503181045:2015/03/18/nfcapd.201503181250 -n 10 -s ip/packets
nfdump filter:
any
Top 10 IP Addr ordered by packets:
Date first seen          Duration Proto           IP Addr    Flows(%)      Packets(%)       Bytes(%)         pps      bps   bpp
2015-01-27 17:44:13.536 4302639.348 any         172.24.2.2    16039( 0.5)  550.8 M(47.6)  669.0 G(55.9)      128    1.2 M  1214
2015-01-27 18:05:10.508 4301378.208 any       172.24.2.100     6073( 0.2)  226.2 M(19.5)  285.7 G(23.9)       52   531287  1262
2015-01-27 17:43:57.588 4302653.688 any       172.23.77.45     1.0 M(30.4)   65.2 M( 5.6)   39.9 G( 3.3)       15    74221   611
2015-03-18 10:15:56.326  9492.706 any      10.195.250.26      211( 0.0)   44.7 M( 3.9)   65.0 G( 5.4)     4708   54.7 M  1453
2015-03-18 10:14:36.934  9611.360 any       172.24.0.10       618( 0.0)   43.0 M( 3.7)   35.6 G( 3.0)     4474   29.7 M   828
2015-03-18 10:15:28.930  9527.426 any       172.24.0.11       248( 0.0)   41.1 M( 3.6)   31.7 G( 2.7)     4316   26.6 M   770
2015-03-18 10:14:04.918  9642.312 any       172.24.0.12       195( 0.0)   40.7 M( 3.5)   25.6 G( 2.1)     4217   21.3 M   630
2015-03-18 10:13:41.338  9665.892 any      172.23.70.243      191( 0.0)   32.7 M( 2.8)   46.4 G( 3.9)     3387   38.4 M  1416
2015-03-18 10:15:56.314  9524.460 any       10.195.250.8      214( 0.0)   30.5 M( 2.6)   42.4 G( 3.5)     3198   35.6 M  1392
2015-03-18 10:15:56.314  9521.946 any      10.195.250.16      202( 0.0)   28.1 M( 2.4)   39.6 G( 3.3)     2946   33.3 M  1410

Summary: total flows: 3302693, total bytes: 1.2 T, total packets: 1.2 G, avg bps: 2.2 M, avg pps: 269, avg bpp: 1032
Time window: 2015-01-27 17:42:34 - 2015-03-18 12:54:53
Total flows processed: 3302693, Blocks skipped: 0, Bytes read: 171758084
Sys: 0.725s flows/second: 4552198.7  Wall: 0.722s flows/second: 4568400.3
```

Рисунок 8. Вывод отчета

Таблица 4. Характеристики, получаемые с помощью NetFlow

IP адрес источника, IP адрес отправителя;	да
TCP/UDP порт источника, TCP/UDP порт отправителя	да
Задержка (мин/макс/сред)	нет
Джиттер	нет
Процент потерянных пакетов	нет
Объем трафика, передаваемый в секунду	да
Количество IP пакетов, передаваемое в секунду	да
Длина IP пакета (мин/макс/**сред**)	частично

Выявление прикладных протоколов	нет

NetFlow изначально создавался как протокол для сбора статистической информации в IP-сетях и он хорошо подходит для учета трафика без его распознавания. Эти данные можно хранить много лет, по необходимости проводя их анализ. Протокол так же подходит для расследования инцидентов ИБ, для планирования и прогнозирования пропускной способности линий связи. Кроме этого NetFlow позволяет проводить мониторинг еще ряда параметров IP пакетов и TCP сегментов не связанных с учетом пользовательского трафика, но позволяющих мониторить качество предоставляемых пользователям услуг (DSCP + ECN).

DPI – Deep Packet Inspection

Deep Packet Inspection (с англ. – *глубокий анализ пакетов*) — технология накопления статистических данных, проверки и фильтрации сетевых пакетов по их содержимому. В отличие от всех предыдущих методов (за исключением некоторых прокси-серверов, заметно уступающих в производительности) Deep Packet Inspection анализирует *не только заголовки пакетов, но и полное содержимое (поля данных)* на уровнях модели TCP/IP со второго и выше, включая уровень приложений.

Приведем пример того, как DPI распознает разные виды трафика прикладного уровня:

– Идентификация BitTorrent

Клиенты BitTorrent соединяются с трекером по протоколу TCP. Для того, чтобы обнаружить среди всего трафика TCP такие пакеты, достаточно проверить, что содержимое данных TCP пакета со второго байта совпадает с «BitTorrent protocol».

– Идентификация HTTP

Для идентификации HTTP протокола достаточно проверить, что пакет является TCP, и содержимое этого TCP пакета начинается с одной из следующих команд: «GET», «POST», «HEAD». Кроме того, после команды должен стоять пробел, а также через некоторый промежуток должен встретиться текст «HTTP/». Если всё это выполняется, то этот пакет несёт в себе HTTP запрос.

Таблица 5. Характеристики, получаемые с помощью NetFlow

IP адрес источника, IP адрес отправителя;	да
TCP/UDP порт источника, TCP/UDP порт отправителя	да
Задержка (мин/макс/сред)	не везде
Джиттер	не везде
Процент потерянных пакетов	не везде
Объем трафика, передаваемый в секунду	да
Количество IP пакетов, передаваемое в секунду	да
Длина IP пакета (мин/макс/сред)	да
Выявление прикладных протоколов	да

DPI это устройство, как правило, специализированное именно на анализе трафика, обладающее широчайшими возможностями как по учету каких-либо данных о трафике, так и по детектированию/блокированию нежелательного контента. Вплоть до контроля утечек критичных файлов за контур внутренней сети (например, базы данных пользователей банка и их счетов). Единственный недостаток – стоимость, отличающаяся в

разы и десятки раз по сравнению с другими методами, и относительная незрелость технологии из-за чего множество функций остается пока еще не реализовано всеми производителями в одном универсальном устройстве.

Одна из вариаций DPI, используемая Российскими спецслужбами – система оперативно розыскных мероприятий, СОРМ (рис. 9). Это система для обеспечения долгосрочного хранения и оперативного доступа к данным об абонентах оператора связи и оказанных услугах связи. Выдаёт структурированную информацию о человеке, номере телефона, звонках, посещенных сайтах, сессиях, использованных прокси и др. по одному какому-либо параметру (например, по ФИО).

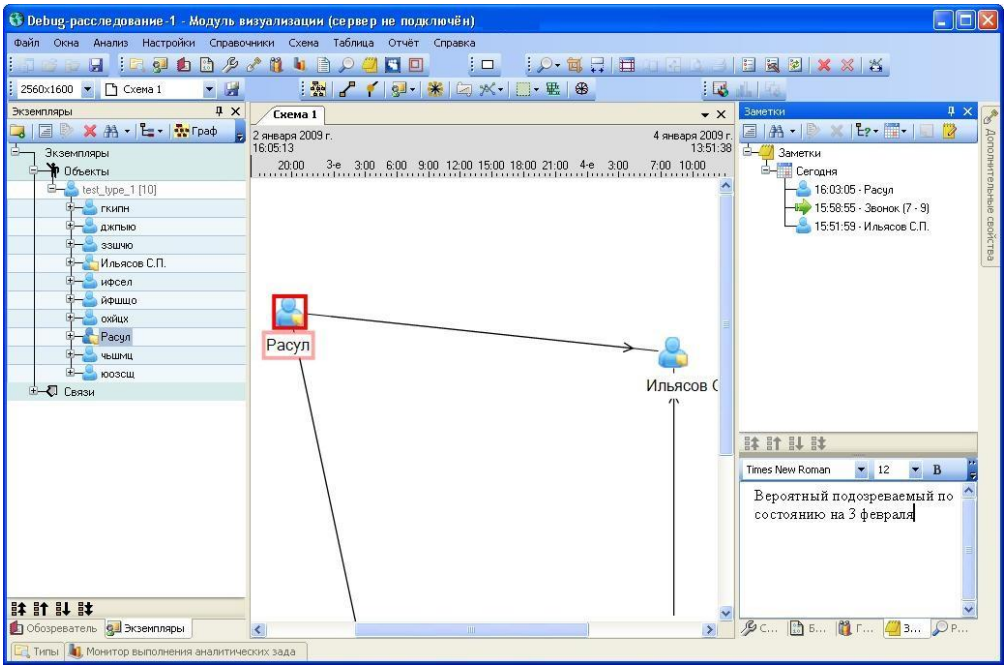


Рисунок 9. Интерфейс управления СОРМ (устаревший)

Вывод

Таким образом, методами, являющимися достаточно информативными и масштабируемыми являются сбор NetFlow, использование proxy-сервер или DPI. Экономически эффективный метод – внедрение proxy-сервера, но при этом нарушается прозрачность в сети. Если на используемом оборудовании есть поддержка протокола NetFlow, то можно без больших финансовых потерь и трудозатрат настроить сбор данных в NetFlow-коллектор, но с данным методом администратор не сможет применять политики к трафику по L7-признакам. Наиболее оптимальным будет установка устройства с поддержкой DPI, но вместе с этим, этот метод является наиболее затратным.

Список литературы

1. Олифер В.Г. Компьютерные сети. Принципы, технологии, протоколы./ В. Г. Олифер, Н.А. Олифер. – М. : Питер, 2011. 944 с.

2. Танненбаум Э. Компьютрные сети. – М. : Питер, 2009. 992 с.

3. Introduction to Cisco IOS NetFlow - A Technical Overview, статья; URL: http://www.cisco.com/c/en/us/products/collateral/ios-nx-os-software/ios-netflow/prod_white_paper0900aecd80406232.html (дата обращения: 04.12.2015)

4. Cisco Systems NetFlow Services Export Version 9, стандарт; URL: https://www.ietf.org/rfc/rfc3954.txt (дата обращения: 04.12.2015)

5. Specification of the IP Flow Information Export (IPFIX) Protocol for the Exchange of Flow Information, стандарт; URL: https://tools.ietf.org/html/rfc7011 (дата обращения: 04.12.2015)

References

1. Olifer V.G. Computer networks. Principles, technologies, protocols./ V. G. Olifer, N.A. Olifer. – Moscow: Piter, 2011. 944 p.

2. Tannenbaum A. Computer networks. – Moscow: Piter, 2009. 992 p.

3. Introduction to Cisco IOS NetFlow - A Technical Overview, article; URL: http://www.cisco.com/c/en/us/products/collateral/ios-nx-os-software/ios-netflow/prod_white_paper0900aecd80406232.html (last visit: 04.12.2015)

4. Cisco Systems NetFlow Services Export Version 9, standard; URL: https://www.ietf.org/rfc/rfc3954.txt (last visit: 04.12.2015)

5. Specification of the IP Flow Information Export (IPFIX) Protocol for the Exchange of Flow Information, standard; URL: https://tools.ietf.org/html/rfc7011 (last visit: 04.12.2015)