АНАЛИЗ И ВИЗУАЛИЗАЦИЯ СЕТЕВОГО ТРАФИКА НА ОСНОВЕ ТЕХНОЛОГИИ ЭКСПОРТА ПОТОКОВ NETFLOW

ANALYSIS AND VISUALIZATION OF NETWORK TRAFFIC BASED ON NETFLOW FLOW EXPORT TECHNOLOGY

P. Zemzerov
S. Suvorov

Summary. NetFlow data provides network operators with detailed information on how it is used. Using this data, you can understand who generates the most traffic, who is the object of DoS attacks, who sends spam, etc. NetFlow is also commonly used to automatically detect anomalies. The problem here is that the amount of NetFlow data collected in a typical backbone network is so large that even the best anomaly detection algorithms do not work either because of the lack of anomaly detection or because there are too many false positives. Another alternative approach to NetFlow data processing is to use visual analytics, that is, to represent large amounts of data in the form of images and animations so that people can detect anomalies during visual inspection. The main goal of this project will be to review existing methods and tools for advanced visualization of NetFlow data, as well as to offer and test a new combination of this data.

Keywords: monitoring systems, telemetry, routers, telecommunications, network security, big data.

Земзеров Павел Андреевич
Московский политехнический университет
dame2509@gmail.com
Суворов Станислав Вадимович
К.э.н., доцент, ФГБОУ ВО «Московский политехнический университет»
ssw1168@mail.ru

Аннотация. NetFlow предоставляют операторам сети, подробную информацию о том, как она используется. С помощью этих данных можно понять, кто генерирует больше всего трафика, кто является объектом DoS-атак, кто рассылает спам и т.д. NetFlow также обычно используется для автоматического обнаружения аномалий. Проблема здесь заключается в том, что объем данных NetFlow, собранных в типичной магистральной сети, столь велик, что даже лучшие алгоритмов обнаружения аномалий не работают либо из-за отсутствия обнаружения аномалий, либо из-за слишком большого количества ложных срабатываний. Другой альтернативный подход к обработке данных NetFlow, состоит в том, чтобы использовать визуальную аналитику, то есть представлять большие объемы данных в виде изображений и анимаций, чтобы люди могли обнаружить аномалии при визуальном осмотре. Основной целью этого проекта будет обзор существующих методов и инструментов для расширенной визуализации данных NetFlow, а также предложение и тестирование новой комбинации этих данных.

Ключевые слова: системы мониторинга, телеметрия, маршрутизаторы, телекоммуникации, сетевая безопасность, большие данные.

Введение

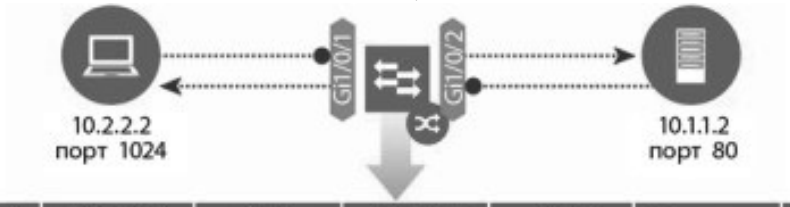
Сетевая безопасность, мониторинг и большие данные становятся основными столпами в сетевой промышленности. Трафик данных растет с угрожающей скоростью и никаких тенденций к его уменьшению нет. Большая часть этого трафика безвредна и не затрагивает интересы систем сетевого мониторинга. Но по мере увеличения объема трафика, становится все труднее отделять вредоносное ПО от безвредного. В настоящее время приемлемый уровень безопасности может быть достигнут с помощью стандарта Cisco NetFlow, с такими инструментами как nfdump [1]. Визуальное представление в таких инструментах ограничено, и не очень интерактивно или интуитивно понятно.

Стандарт NetFlow [2] создает среду, которая включает в себя инструменты, позволяющие понять, кто, что, когда, где и как передает сетевой трафик. NetFlow облегчает администраторам оптимальное использование сети. Можно определить источник и назначение трафика и использовать эту информацию для выявления, например, спама или DDoS-атак.

Каждый пакет, который пересылается в маршрутизаторе / коммутаторе, проверяется на набор атрибутов пакета Интернет-протокола (IP). С помощью этих атрибутов можно определить, является ли пакет уникальным или похожим на другие пакеты.

NetFlow использует следующие атрибуты:
1. IP-адрес источника
2. IP-адрес назначения адресата
3. Порт источник
4. Порт назначения
5. Протокол 3 уровня
6. Класс обслуживания
7. Интерфейс маршрутизатора / коммутатора

Чтобы сгруппировать пакеты в поток, сравнивается IP-адрес источника/назначения, порты источника/назначения, интерфейс протокола и класс обслуживания.



Кеш потока

Время начала	Интерфейс	IP-адрес отправителя	Порт отправителя	IP-адрес получателя	Порт получателя	Протокол	Отправлено пакетов	Отправлено байтов
10:20:12.221	Gig 1/0/1	10.2.2.2	1024	10.1.1.2	80	TCP	5	1025
10:20:12.871	Gig 1/0/1	10.2.2.2	80	10.2.2.2	1024	TCP	17	28712

Рис. 1. Создание потока в кеше NetFlow [3]

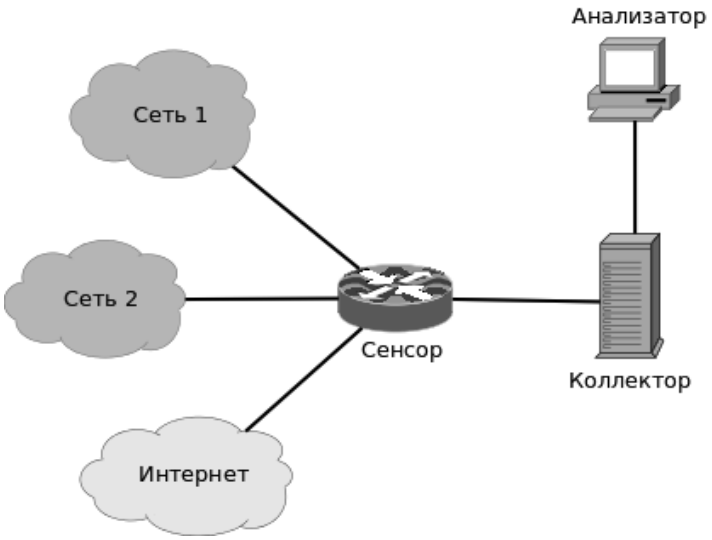


Рис. 2. Простая архитектура NetFlow

Затем пакеты и байты подсчитываются. Этот метод является масштабируемым, потому что большой объем сетевой информации конденсируется в базе данных с информацией NetFlow, называемой кэш-памятью NetFlow, рисунок 1.

Когда кэш NetFlow создан, его можно использовать для понимания поведения сети. Различные атрибуты генерируют разные знания об определенной сети, и в совокупности они могут нарисовать детальную картину того, как работает сеть. Например:

- Интерфейс сообщает, как сетевое устройство использует трафик.
- Адрес отправителя позволяет понять, кто инициирует трафик
- Адрес получателя сообщает, кто получает трафик
- Порты характеризуют приложение, использующее трафик
- Протокол проверяет приоритет трафика.
- Переданные пакеты и байты показывают объем трафика.

Дополнительная информация, добавляемая к потоку, включает в себя:

- Метки времени потока для понимания срока службы потока. Метки полезны для вычисления пакетов и байтов в секунду
- IP-адреса следующего перехода, включая маршрутизацию по протоколу пограничного шлюза (BGP). Автономные системы (AS)
- Маску подсети для адресов источника и назначения для вычисления префиксов
- Флаги для проверки коннекта по протоколу управления передачей (TPC)

Основные компоненты NetFlow:

- Сенсор: объединяет пакеты в потоки и экспортирует записи потока в одну единицу или более

```
router#show ip cache flow
IP packet size distribution (90784136 total packets):
  1-32  64  96 128 160 192 224 256 288 320 352 384 416 448 480
  .000 .698 .011 .001 .004 .005 .000 .004 .000 .000 .003 .000 .000 .000 .000
  512 544 576 1024 1536 2048 2560 3072 3584 4096 4608
  .000 .001 .256 .000 .010 .000 .000 .000 .000 .000 .000
IP Flow Switching Cache, 4456704 bytes
 1885 active, 63651 inactive, 59960004 added
 129803821 ager polls, 0 flow alloc failures
 Active flows timeout in 30 minutes
 Inactive flows timeout in 15 seconds
IP Sub Flow Cache, 402056 bytes
 0 active, 16384 inactive, 0 added, 0 added to flow
 0 alloc failures, 0 force free
 1 chunk, 1 chunk added
 last clearing of statistics never
Protocol         Total    Flows   Packets Bytes  Packets Active(Sec) Idle(Sec)
--------         Flows     /Sec     /Flow  /Pkt     /Sec     /Flow     /Flow
TCP-Telnet    11393421      2.8         1    48      3.1       0.0       1.4
TCP-FTP           236      0.0        12    66      0.0       1.8       4.8
TCP-FTPD           21      0.0     13726  1294      0.0      18.4       4.1
TCP-WWW         22282      0.0        21  1020      0.1       4.1       7.3
TCP-X             719      0.0         1    40      0.0       0.0       1.3
TCP-BGP             1      0.0         1    40      0.0       0.0      15.0
TCP-Frag        70399      0.0         1   688      0.0       0.0      22.7
TCP-other     47861004     11.8         1   211     18.9       0.0       1.3
UDP-DNS           582      0.0         4    73      0.0       3.4      15.4
UDP-NTP         287252      0.0         1    76      0.0       0.0      15.5
UDP-other       310347      0.0         2   230      0.1       0.6      15.9
ICMP             11674      0.0         3    61      0.0      19.8      15.5
IPv6INIP            15      0.0         1  1132      0.0       0.0      15.4
GRE                  4      0.0         1    48      0.0       0.0      15.3
Total:        59957957     14.8         1   196     22.5       0.0       1.5
SrcIf         SrcIPaddress    DstIf         DstIPaddress    Pr SrcP DstP  Pkts
Gi0/0         192.168.10.201  Gi0/1         192.168.60.102  06 0984 0050     1
Gi0/0         192.168.11.54   Gi0/1         192.168.60.158  06 0911 0035     3
Gi0/1         192.168.150.60  Gi0/0         10.89.16.226    06 0016 12CA     1
Gi0/0         192.168.10.17   Gi0/1         192.168.60.97   11 0B89 0050     1
Gi0/0         10.88.226.1     Gi0/1         192.168.202.22  11 007B 007B     1
Gi0/0         192.168.12.185  Gi0/1         192.168.60.239  11 0BD7 0050     1
Gi0/0         10.89.16.226    Gi0/1         192.168.150.60  06 12CA 0016     1
router#
```

Рис. 3. Как выглядит DoS атака в необработанном формате [6]

коллекторов потока. Обычно это L3-коммутатор или маршрутизатор, хотя можно использовать и отдельно стоящие сенсоры, получающие данные путём зеркалирования порта коммутатора.

- Коллектор (сборщик потока): отвечает за прием, хранение и предварительную обработку данных потока, полученных от экспортера потока.
- Анализатор: приложение, которое анализирует полученные данные потока в различных контекстах, формирует пригодные для чтения человеком отчёты (часто в виде графиков)

Nfdump собирает и обрабатывает данные NetFlow в командной строке. Он хранит данные NetFlow во временных файлах. Файлы являются двоичными, и это дает возможность либо вернуть выходные данные из nfdump в той же двоичной форме, либо в виде читаемого текста. У nfdump есть четыре формата вывода: raw, line, long и extended. Задача представления адресов Интернет-протокола версии 6 (IPv6) решается путем сокращения их в обычном выводе. [4]

Кстати существует несколько версий протокола NetFlow:

NetFlow v5 ограничивается потоками IPv4. Выбор полей, которые можно экспортировать с использованием этой версии, тоже ограничен.

NetFlow v9 имеет ряд преимуществ перед предшествующими форматами. Шаблон v9 позволяет определять, что помещается в запись. Уменьшив детализацию, можно поместить в дейтаграмму больше потоков.

IPFIX — это стандарт, рассматривающий поток как любое число пакетов, наблюдаемых в определенном интервале времени и имеющих ряд общих свойств (например, один и тот же отправитель, один и тот же протокол). IPFIX позволяет отправляющему устройству включать в сообщения предопределенные типы данных, которые пользователь может задать при помощи специальных шаблонов. [3]

Сетевые атаки. За последнее десятилетие важность защиты от атак на большие компьютерные системы сильно возросла. В 2004 году на семинаре ACM по визуализации и интеллектуальному анализу данных для компьютерной безопасности была представлена технология NVisionIP: визуализация сетевых потоков для обеспечения ситуационной осведомленности безопасности [5]. Это был один из первых инструментов для визуализации данных NetFlow. Визуализация основывалась либо на количестве переданных байтов, либо на количестве потоков к или от хостов в сети. Из-за его ограничений, он не был проверен в ходе этой работы. Интерфейс NVisionIP можно увидеть на рисунке 7.

В [5] ACM обсуждают использование NVisionIP для борьбы с различными проблемами безопасности. Большинство подобных атак, описанных в этой работе, актуальны сегодня, хотя в сегодняшних огромных объемах данных их сложнее обнаружить.

- Червь: одна из основных функций безопасности, которую можно обнаружить. Черви обычно распространяются путем поиска других хозяев. Отфильтровывая хосты, передающие много потоков через один порт назначения, можно легко увидеть, какие машины заражены, и их следует отключить.
- Скомпрометированные системы: если хост скомпрометирован, злоумышленник может установить вредоносное ПО, позволяющее злоумышленнику контролировать компьютер. После этого злоумышленник может превратить хост в файловый сервер. Обнаружив большие объемы трафика на определенных портах, можно обнаружить такую атаку.
- Сканирование портов: когда на определенном хосте используется большое количество портов, оно легко идентифицируется NVisionIP.
- Dos-атаки: отличительной моделью поведения Dos-атак является резкое увеличение исходящего трафика со стороны атакующего хоста. Если же сам хост атакован, тот же шаблон виден в увеличенном объеме получаемого трафика. Таким образом, пики в трафике не обязательно являются атакой, но могут быть результатом нового выпуска или резервного копирования. Чтобы иметь возможность идентифицировать DDoS-атаку, можно взглянуть на нее с двух сторон. Находя кого-то, кто атакует, или кого-то, кто подвергается нападению. Рассмотрим второй сценарий. Если кто-то является целью DDoS, появится внезапный пик входящего трафика. После дальнейшего изучения можно будет найти аналогичное поведение сети среди предыдущих данных, чтобы найти шаблон, который можно раскрыть при реальной атаке.

Есть несколько потоков для UDP-порта 80 (шестнадцатеричное значение 0050). Кроме того, существуют также потоки для TCP-порта 53 (шестнадцатеричное значение 0035) и TCP-порта 80 (шестнадцатеричное значение 0050). Пакеты в этих потоках могут быть подделаны и могут указывать на попытку выполнить эти атаки. Рекомендуется сравнить потоки для TCP-порта 53 (шестнадцатеричное значение 0035) и TCP-порта 80 (шестнадцатеричное значение 0050) с обычными базовыми показателями, чтобы определить, идет ли атака. [6]

Предполагается, что данные уже были обработаны, прежде чем они становятся доступными. Мне предоставили два месяца анонимных данных от BITRACE, чтобы ознакомиться с NetFlow и иметь возможность использовать реальные данные для визуализаций. Это анонимные данные за январь 2012 года от сборщиков Trondheim и Oslo NetFlow. Данные с этих коллекторов отбираются в соотношении 1/100 или 1/1000. Выборка может привести к некоторым отклонениям в поведении сетей, но из-за характера атак, исследованных в этой работе, выборка не должна быть препятствием. Атаки, в которых критичен только конкретный поток, такой коэффициент выборки мог бы исключить эти пакеты, и атака прошла бы незамеченной.

Было решено сосредоточиться на самых популярных IP-адресах и портах назначения в разные периоды времени. Подробно рассмотрим почасовой просмотр каждого IP и порта. Выбирая адресный спектр в качестве основного, необходимо найти способ представления всего спектра IPv4. Это одна проблема, и когда дело доходит до IPv6, это становится практически невозможным из-за диапазона адресов. Это значит, что придется использовать предварительную обработку, чтобы выделить IP-адреса, которые стоит рассмотреть поближе. В данных,

```
[eeglarse@iou2:/data/netflow$ find . -printf '%s %p\n'|sort -nr|head
14838848 ./oslo_gw/2012/01/18/nfcapd.201201181325
14729440 ./oslo_gw/2012/01/18/nfcapd.201201181335
14729284 ./oslo_gw/2012/01/18/nfcapd.201201181310
14720548 ./oslo_gw/2012/01/18/nfcapd.201201181330
14687944 ./oslo_gw/2012/01/18/nfcapd.201201181315
14651908 ./oslo_gw/2012/01/18/nfcapd.201201181340
14566420 ./oslo_gw/2012/01/18/nfcapd.201201181320
14563196 ./oslo_gw/2012/01/18/nfcapd.201201181305
14508804 ./oslo_gw/2012/01/18/nfcapd.201201181345
14472664 ./oslo_gw/2012/01/18/nfcapd.201201181300
```

Рис. 4. Файлы с наибольшим количеством потоков из предоставленных файлов.

```
[eeglarse@iou2:/data/netflow$ find . -printf '%s %p\n'|sort -nr|tail -100
849408 ./trd_gw1/2012/01/01/nfcapd.201201010920
848160 ./trd_gw1/2012/01/01/nfcapd.201201010745
842856 ./trd_gw1/2012/01/01/nfcapd.201201010725
834212 ./trd_gw1/2012/01/01/nfcapd.201201010655
832340 ./trd_gw1/2012/01/01/nfcapd.201201010640
830364 ./trd_gw1/2012/01/01/nfcapd.201201010555
828856 ./trd_gw1/2012/01/01/nfcapd.201201010845
821940 ./trd_gw1/2012/01/01/nfcapd.201201010900
816012 ./trd_gw1/2012/01/01/nfcapd.201201010905
804624 ./trd_gw1/2012/01/01/nfcapd.201201010735
802596 ./trd_gw1/2012/01/01/nfcapd.201201010545
799164 ./trd_gw1/2012/01/01/nfcapd.201201010635
780600 ./trd_gw1/2012/01/01/nfcapd.201201010650
772644 ./trd_gw1/2012/01/01/nfcapd.201201010610
760424 ./trd_gw1/2012/01/01/nfcapd.201201010705
758864 ./trd_gw1/2012/01/01/nfcapd.201201010720
```

Рис. 5. Файлы с наименьшим количеством потоков из предоставленных файлов

```
[eeglarse@iou2:~$ cat test_180112.csv |cut -f 5 -d ',' |sort|uniq -c|sort|tail
  17762 162.185.32.85
  19878 161.222.192.123
  21506 191.220.233.80
  23995 161.223.1.164
  37704 161.223.1.108
  39759 159.152.145.176
  49316 161.223.1.142
  51467 190.49.180.97
  61424 161.223.1.106
 120976 192.239.62.2
```

Рис. 6. Первая десятка использованных адресов назначения за период 1300–1400, 18 января

```
eeglarse@iou2:~$ cat test_180112.csv |cut -f 4 -d ',' |sort|uniq -c|sort|tail -10
  18502 161.222.192.123
  18557 191.220.233.80
  19338 162.185.32.85
  29367 161.223.1.164
  46376 192.239.62.2
  47139 190.49.180.97
  47844 161.223.1.108
  50509 161.223.1.142
  77527 159.152.145.176
  83184 161.223.1.106
```

Рис. 7. Первая десятка адресов источника за период 1300–1400, 18 января

```
eeglarse@iou2:~$ nfdump -R /data/netflow/oslo_gw/2012/01/18/nfcapd.201201180000:nfcapd.201201182355 -n 10 -s dstport 'dst ip 192.239.62.2'
Top 10 Dst Port ordered by flows:
Date first seen          Duration Proto          Dst Port    Flows(%)      Packets(%)       Bytes(%)          pps      bps   bpp
2012-01-18 08:05:55.302 21426.006 any              19424      224( 0.0)       284( 0.0)    284614( 0.0)        0      106  1002
2012-01-18 09:08:59.910 23270.338 any               8981      216( 0.0)       287( 0.0)    298504( 0.0)        0      102  1040
2012-01-18 08:07:44.051 31093.199 any              51750      207( 0.0)       223( 0.0)    252365( 0.0)        0       64  1131
2012-01-18 09:11:42.207 39490.631 any              40293      156( 0.0)       219( 0.0)    190328( 0.0)        0       38   869
2012-01-18 07:57:48.578 31937.555 any              31019      153( 0.0)       601( 0.0)    838518( 0.0)        0      210  1395
2012-01-18 08:44:26.768 25405.425 any              32586      138( 0.0)       398( 0.0)    534649( 0.0)        0      168  1343
2012-01-18 10:07:20.801 28733.933 any              35708      113( 0.0)       312( 0.0)    419602( 0.0)        0      116  1344
2012-01-18 11:06:40.021 14417.571 any               7211      113( 0.0)       357( 0.0)    489458( 0.0)        0      271  1371
2012-01-18 09:00:24.646 15560.048 any              65267      107( 0.0)       213( 0.0)    292432( 0.0)        0      150  1372
2012-01-18 07:47:39.195 29887.494 any              55562      106( 0.0)       558( 0.0)    770732( 0.0)        0      206  1381

Summary: total flows: 830383, total bytes: 1.9 G, total packets: 1.6 M, avg bps: 178956, avg pps: 18, avg bpp: 1224
Time window: <unknown>
Total flows processed: 44250006, Blocks skipped: 0, Bytes read: 2301034764
Sys: 2.840s flows/second: 15579254.6 Wall: 2.843s flows/second: 15559172.3
```

Рис. 8. Топ-10 самых популярных портов назначения и количество потоков

```
eeglarse@iou2:~$ nfdump -R /data/netflow/oslo_gw/2012/01/18/nfcapd.201201180000:nfcapd.201201182355 -n 10 -s dstport 'dst ip 192.239.62.2'
Top 10 Dst Port ordered by flows:
Date first seen          Duration Proto          Dst Port    Flows(%)      Packets(%)       Bytes(%)          pps      bps   bpp
2012-01-18 08:05:55.302 21426.006 any              19424      224( 0.0)       284( 0.0)    284614( 0.0)        0      106  1002
2012-01-18 09:08:59.910 23270.338 any               8981      216( 0.0)       287( 0.0)    298504( 0.0)        0      102  1040
2012-01-18 08:07:44.051 31093.199 any              51750      207( 0.0)       223( 0.0)    252365( 0.0)        0       64  1131
2012-01-18 09:11:42.207 39490.631 any              40293      156( 0.0)       219( 0.0)    190328( 0.0)        0       38   869
2012-01-18 07:57:48.578 31937.555 any              31019      153( 0.0)       601( 0.0)    838518( 0.0)        0      210  1395
2012-01-18 08:44:26.768 25405.425 any              32586      138( 0.0)       398( 0.0)    534649( 0.0)        0      168  1343
2012-01-18 10:07:20.801 28733.933 any              35708      113( 0.0)       312( 0.0)    419602( 0.0)        0      116  1344
2012-01-18 11:06:40.021 14417.571 any               7211      113( 0.0)       357( 0.0)    489458( 0.0)        0      271  1371
2012-01-18 09:00:24.646 15560.048 any              65267      107( 0.0)       213( 0.0)    292432( 0.0)        0      150  1372
2012-01-18 07:47:39.195 29887.494 any              55562      106( 0.0)       558( 0.0)    770732( 0.0)        0      206  1381

Summary: total flows: 830383, total bytes: 1.9 G, total packets: 1.6 M, avg bps: 178956, avg pps: 18, avg bpp: 1224
Time window: <unknown>
Total flows processed: 44250006, Blocks skipped: 0, Bytes read: 2301034764
Sys: 2.840s flows/second: 15579254.6 Wall: 2.843s flows/second: 15559172.3
```

Рис. 9. Топ 10 самых популярных исходных портов и количество потоков

```
eeglarse@iou2:~$ nfdump -R /data/netflow/oslo_gw/2012/01/18/nfcapd.201201180000:nfcapd.201201182355 -n 10 -s srcport 'dst ip 192.23
9.62.2'
Top 10 Src Port ordered by flows:
Date first seen          Duration Proto          Src Port    Flows(%)      Packets(%)       Bytes(%)          pps      bps   bpp
2012-01-17 23:58:26.575 86431.371 any                 80   753701(90.8)     1.5 M(93.6)     1.8 G(95.7)       17   171173  1251
2012-01-17 23:59:47.399 86318.714 any                443    70569( 8.5)     86400( 5.5)    67.7 M( 3.5)        1     6269   782
2012-01-18 00:00:04.416 86245.775 any                 53     2068( 0.2)      2068( 0.1)    304428( 0.0)        0       28   147
2012-01-18 08:06:13.613 38095.737 any               8080     1283( 0.2)      2158( 0.1)     2.9 M( 0.2)        0      616  1360
2012-01-18 08:16:18.036 37528.850 any               8182      492( 0.1)      5510( 0.3)     7.8 M( 0.4)        0     1666  1418
2012-01-18 08:14:50.332 37657.316 any                364      238( 0.0)       937( 0.1)     1.3 M( 0.1)        0      282  1420
2012-01-18 08:07:44.051 31093.199 any               8088      222( 0.0)       238( 0.0)    256549( 0.0)        0       66  1077
2012-01-18 08:05:55.302 17131.174 any               8004      218( 0.0)       276( 0.0)    274566( 0.0)        0      128   994
2012-01-18 10:51:48.542 15511.827 any               8003      212( 0.0)       281( 0.0)    289984( 0.0)        0      149  1031
2012-01-18 08:51:01.496 26206.793 any               8000      172( 0.0)       231( 0.0)    165360( 0.0)        0       50   715

Summary: total flows: 830383, total bytes: 1.9 G, total packets: 1.6 M, avg bps: 178956, avg pps: 18, avg bpp: 1224
Time window: <unknown>
Total flows processed: 44250006, Blocks skipped: 0, Bytes read: 2301034764
Sys: 2.858s flows/second: 15479569.6 Wall: 2.856s flows/second: 15489030.1
```

Рис. 10. Топ-10 самых популярных исходных портов и количество потоков, отправляемых на IP-адрес 192.239.62.2

предоставленных BITRACE, можно, например, перечислить 10 лучших файлов по размеру, что означает временные интервалы с наибольшим количеством потоков. Этот запрос предоставил результаты на рисунке 4, когда был выполнен с предоставленными данными.

Из этой простой предварительной обработки легко увидеть, что в период между 1300–1400 и далее 18 января был четкий пик потоков. Если мы сравним их с файлами с наименьшим количеством потоков, как показано на рисунке 5, мы увидим большой промежуток между файлами с наибольшим и наименьшим количеством потоков. Создаем файл.csv, содержащий рассматриваемый час. С помощью этого файла можно найти наиболее часто используемые IP-адреса назначения, показанные на рисунке 6. Один конкретный адрес четко отделен от других. Большие значения могут быть DDoS-атакой или другими нарушениями, но это необязательно, возможны редкие аномалии. Если мы посмотрим на список верхних IP-адресов отправляющих пакетов, то на рисунке 6 мы увидим, что тот же IP-адрес 192.239.62.2, происходит и здесь.

Для дальнейшего изучения активности на одном определенном IP-адресе можно посмотреть на порты, как источника, так и назначения. В случае упомянутого IP-адреса потоки широко распределены по тысячам портов назначения на рисунке 9. Если мы посмотрим

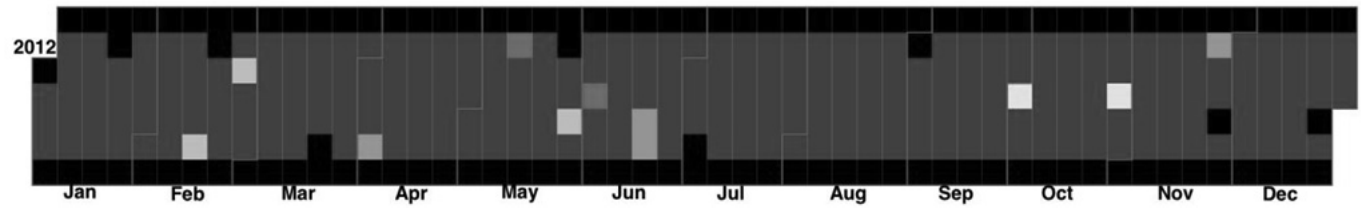


Рис. 11. График статистики по трафику за год.

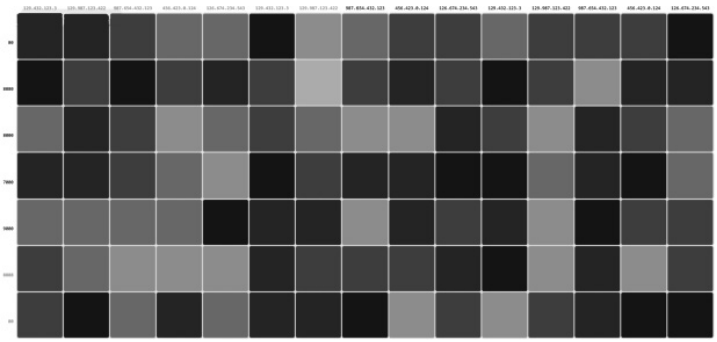


Рис. 12.: Статистика по комбинациям IP и портов.

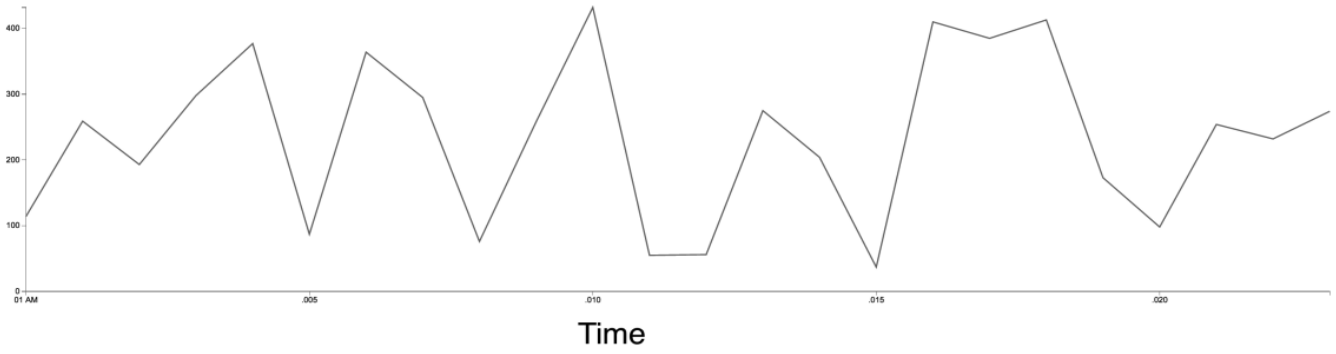


Рис. 13.: 24-часовой график, показывающий поведение на одном IP и комбинации портов

на порты назначения на рисунке 10, мы отметим, что 90.8% трафика исходит из порта 80, это означает, что это, вероятно, кто-то загружает большой файл на наш IP-адрес.

Из пакетов NetFlow также можно посмотреть временной спектр. При огромном количестве IP-адресов этот спектр не подходит для представления данных для обнаружения конкретных атак и т.д. Однако его можно использовать для мониторинга объемов трафика с течением времени или того, какие порты используются в определенные моменты времени и т.д.

Чтобы сократить время загрузки во внешнем интерфейсе и обеспечить максимально возможную понятность данных, их следует разбить мелкие и многочисленные файлы. С помощью этих сценариев мы выполняем расширенную предварительную обработку значительных объемов данных, чтобы вывести желаемый результат и наилучшим образом найти отклонения с помощью визуализации.

Первый визуальный элемент, который у нас есть, это обзор, который в данном случае показывает целый год, аккуратно разделенный на месяцы, недели и дни, как показано на рисунке 11. Цель этого состоит в том, чтобы иметь возможность быстро распознавать шаблоны в данных, которые соответствуют регулярным действиям в качестве резервного копирования и т.д., например, еженедельное резервное копирование будет создавать аналогичные уровни использования сети в определенное время каждую неделю.

Таблица 1. Плюсы и минусы визуального решения.

Визуальное решение	
Плюсы • Интерфейс интуитивно понятен и более прост в использовании. • Не требует специальных навыков, например, как для работы с nfdump и т.д. • Визуальная интерпретация данных быстрее и проще для человеческого понимания. • Шаблоны и другие аберрации, более заметны, нежели в текстовом решении. • Отображаются только нужные данные, пользователь не перегружается лишней информацией.	**Минусы** • Визуальное решение может ограничивать возможности, существующие в решении на основе командной строки. Если вдаваться в детализацию данных, визуальное решение может быть ограничено определенными параметрами и не будет отображать общую картину.

Таблица 2. Плюсы и минусы решения командной строки.

Решение для командной строки	
Плюсы • Решение для командной строки содержит широкий спектр очень специфических команд, как, например, в nfdump • Предоставляет возможность выполнять очень четкие поиски.	**Минусы** • Отнимает много времени и более сложен в освоении. • Шаблоны не легко распознать при использовании командной строки,

IP-адреса и порты. На каждый день существуют миллионы различных комбинаций IP-адресов и портов, которые передают потоки между собой. Благодаря предварительной обработке можно определить, какие IP-адреса являются наиболее популярными каждый день, и, таким образом, найти порты, которые IP-адреса используют чаще всего. Пример в рис. 12 визуализирует количество потоков для каждой из этих комбинаций через тепловую карту. Тепловая карта различает значения в цветовом диапазоне на основе самых высоких значений в наборе данных, то есть чем больше значение, тем темнее цвет.

24-часовой график. Когда IP-адрес и порт уточняются для определенного дня, строим его 24-часовой график и количество потоков в каждый час для выбранного IP-адреса и порта.

Плюсы и минусы визуального и текстового решения.

Оба решения служат своей цели по-разному. Некоторые из минусов каждого из них хорошо дополняются другими плюсами.

Как показывает практика, nfdump способен углубляться и детализировать данные несколько лучше, чем любое визуальное решение. И несмотря на значительный потенциал, визуальное решение не способно показать очень подробную информацию. Но используемое вместе с решением командной строки, визуальное решение избавляет от необходимости выполнить несколько команд nfdump, требующих больших ресурсов, позволяет быстро раскрыть информацию, которая в противном случае потребовала бы много времени для поиска с помощью nfdump.

Как часть большей системы, визуальное решение имеет потенциал. Но его минус в разнообразии, требуется много визуальных элементов, чтобы покрыть все возможные атаки. В целом ясно, что визуальное решение вносит большой вклад в мониторинг сети, но сталкивается с проблемами, которые решаются полностью, потому что требуемый уровень детализации узок и часто ограничен конкретными событиями. Оба решения работают вместе и дополняют друг друга. Они хорошо работают вместе и предоставляют полную функциональность там, где отсутствует другой. Так же визуальное представление не только представляет собой полезный инструмент в аспекте безопасности сетевого мониторинга, но также представляет статистику использования сети.

В этой работе была представлена теория, лежащая в основе Cisco NetFlow и базовой теории визуализации. Это лишь начальный этап в разработке и тестировании простого инструмента для взаимодействия с данными NetFlow для обнаружения необычного поведения в сети, выявляя некоторые аномалии. Потенциал такого решения очевиден. Дальнейший анализ того, какая информация должна быть включена в визуальное представление, может дать решение, способное обнаружить аномалии, не обнаруживаемые современными системами. Показывая, как визуальные элементы могут улучшить взаимодействие людей с данными NetFlow, я надеюсь, что потенциал использования визуальных инструментов

в сочетании с человеческим осмотром и опытом будет усилен. Проведенные исследования не подтвердили, что данные, предоставленные BITRACE, содержали какие-либо атаки, но показали, что при визуальном представлении данных обнаруживаются сходные паттерны, обнаруженные с помощью текущего решения nfdump.

ЛИТЕРАТУРА

1. Установка и настройка nfdump + nfsen для сбора NetFlow [Электронный ресурс] URL https://salatpower.ru/?p=49 (дата обращения: 26.09.2019).
2. Александр Конюхов NetFlow, Cisco и мониторинг трафика. [Электронный ресурс] URL https://habr.com/ru/post/175359/ (дата обращения: 26.09.2019).
3. Джереми Редьярд «Безопасность сети и Netflow» 10.09.2018 [Электронный ресурс] URL: https://www.osp.ru/lan/2018/09/13054476/ (дата обращения: 26.09.2019).
4. Ловим NetFlow при помощи Nfdump и Nfsen [Электронный ресурс] URL: https://voipnotes.ru/lovim-netflow-pri-pomoshi-nfdump-i-nfsen/ (дата обращения: 26.09.2019).
5. Киран Лаккараю «NVisionIP: NetFlow визуализация состояния системы для обеспечения безопасности ситуационной осведомленности» Конференция: семинар по визуализации и интеллектуальному анализу данных для компьютерной безопасности (VizSEC / DMSEC2004), 29 октября 2004 г., Вашингтон, округ Колумбия, США [Электронный ресурс] URL: https://www.researchgate.net/profile/William_Yurcik/publication/221325930_NVisionIP_NetFlow_visualizations_of_system_state_for_security_situational_awareness/links/02bfe50d2367f0dc71000000/NVisionIP-NetFlow-visualizations-of-system-state-for-security-situational-awareness.pdf (дата обращения: 26.09.2019).
6. «Руководство Cisco по усилению защиты устройств Cisco IOS», 2017 г. [Электронный ресурс] URL: https://www.cisco.com/c/ru_ru/support/docs/ip/access-lists/13608–21.html (дата обращения: 26.09.2019).

Журнал «Современная наука: актуальные проблемы теории и практики»

Московский политехнический университет