**Министерство науки и высшего образования Российской Федерации**

**ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ**

**«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ ИТМО»**

**Факультет безопасности информационных технологий**

**Дисциплина:**

**«Управление мобильными устройствами»**

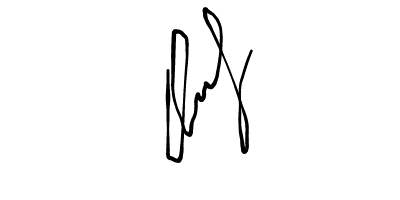
**ОТЧЕТ ПО ЛАБОРАТОРНОЙ РАБОТЕ №2**

**«****Обработка и тарификация трафика NetFlow»**

**Выполнила:**

Студентка гр. N3349

Серова Ольга Евгеньевна



14.04.2020

**Проверил:**

Федоров Иван Романович

Санкт-Петербург

2020 г.

**Цель работы:** обработка трафика NetFlow v5 из данного файла и реализация простейшего правила тарификации этого трафика.

**Задачи:**

1. Привести данный файл в читабельный вид (например с помощью утилиты nfdump)   
   nfdump -r nfcapd.202002251200;
2. Сформировать собственный файл для тарификации любого формата, с которым удобно работать (в соответствии с вариантом работы);
3. Построить график зависимости объема трафика от времени;
4. Протарифицировать трафик в соответствии с вариантом задания.

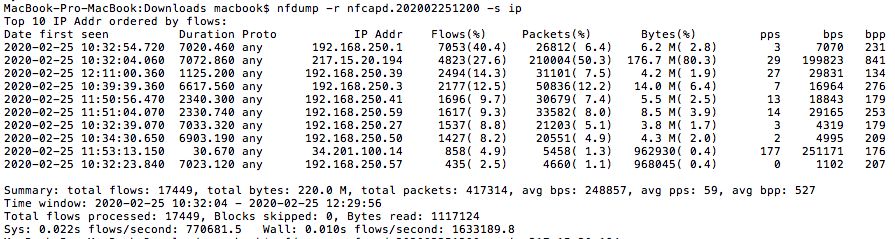
**Ход работы**

Правила тарификации услуг “Интернет”:  
*X = Q \* k*,   
где *X* - итоговая стоимость, *Q* - общий объем трафика NetFlow за отчетный период, *k* - множитель тарифного плана (у каждого варианта свой).

Для того, чтобы привести данные из файла в читабельный вид необходимо выполнить команду:

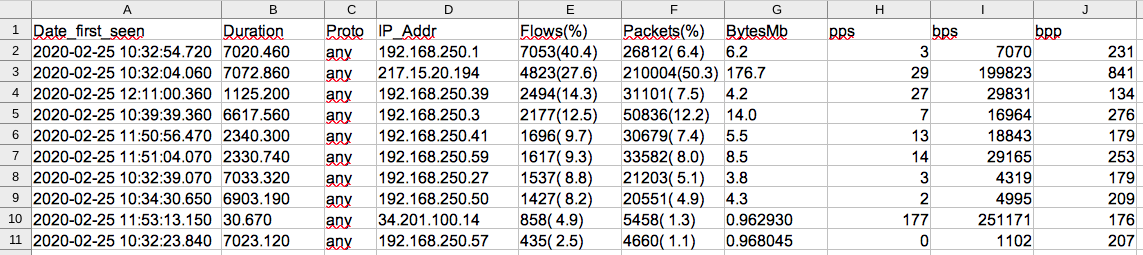
nfdump -r nfcapd.202002251200 -s ip

данная команда предназначена для анализа трафика и выдачи статистики. В нашем случает статистика будет составлена для ip адресов (рисунок 1). В статистику входят такие показателя, как Date first seen, Duration, Proto, IP Addr, Flows(%), Packets(%), Bytes(%), pps, bps, bpp.

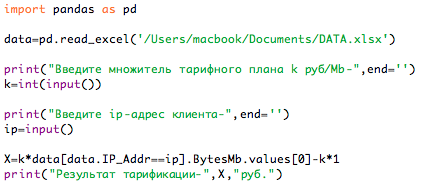
*Рис. 1 «Результат выполнения команды nfdump -r nfcapd.202002251200 -s ip»*

Также стоит отметить то, что данная статистика собирает весь объем трафика по ip адресам, не зависимо от того, был ли рассматриваемый ip адрес адресом приемной для трафика стороны или же адресом отправителя этого трафика.

Полученные данные переносим из терминала в таблицу ***DATA.xlsx*** (рисунок 2) вручную, так как команда *nfdump -r nfcapd.202002251200 -s ip* с ключом *-о cvs* выполняет некорректную конвертацию данных. Стоит также привести все значения столбца Bytes к единому формату представления данные, иными словами в записи каждого отдельного значения не должны присутствовать буквы, их следует заменить умножением на 10 в соответствующей этой буквенной приставке степени.

*Рис.2 «Перенесенные в таблицу данные»*

После того, как файл для тарификации в виде таблицы формата xlsx был сформирован, он был обработан программным модулем листинг которого приведен на рисунке 3:

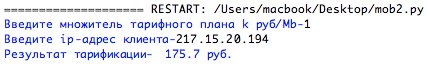
*Рис.3 «Листинг программного модуля* ***mob2.py****»*

Как видно из кода программы для ее реализации так же, как и в первой лабораторной работе был выбран высокоуровневый язык программирования Python3 и библиотека Pandas, предназначенная для анализа и обработки данных. Такой выбор был сделан мной ввиду простоты работы как с самим Python, так и с библиотекой Pandas.

**Вариант 2**

Протарифицировать абонента с IP-адресом 217.15.20.194   
с коэффициентом k: 1руб/Мб, первая 1000Мб бесплатно

Результат тарификации приведен на рисунке 4:

*Рис. 4 «Результата выполнения тарификации клиента с IP-адресом 217.15.20.194 по условиям тарификации»*

Для построения графика зависимости объема трафика от времени необходимо отфильтровать данные NetFlow из файла. Для этого необходимо выполнить команды:

nfdump -r nfcapd.202002251200 >mobile2.txt

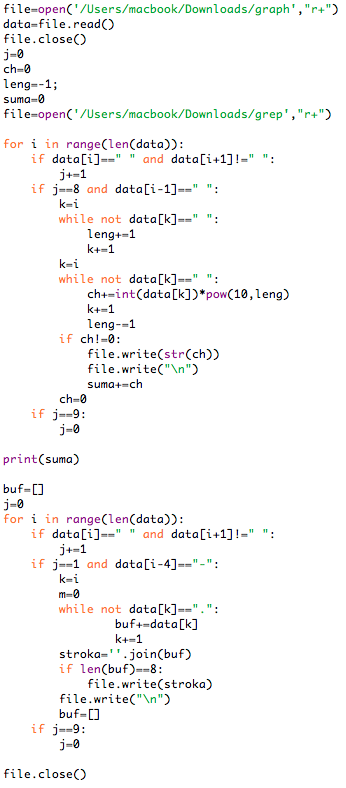
grep -bn 217.15.20.194 mobile2.txt >graph

Первая команда просто сохраняет полученные с помощью утилиты nfdump данные в файл ***mobile2.txt***, а вторая команда находит все строки, в которых встречается заданный условием варианта ip адрес, и сохраняет их в бинарный файл без расширения ***graph***.

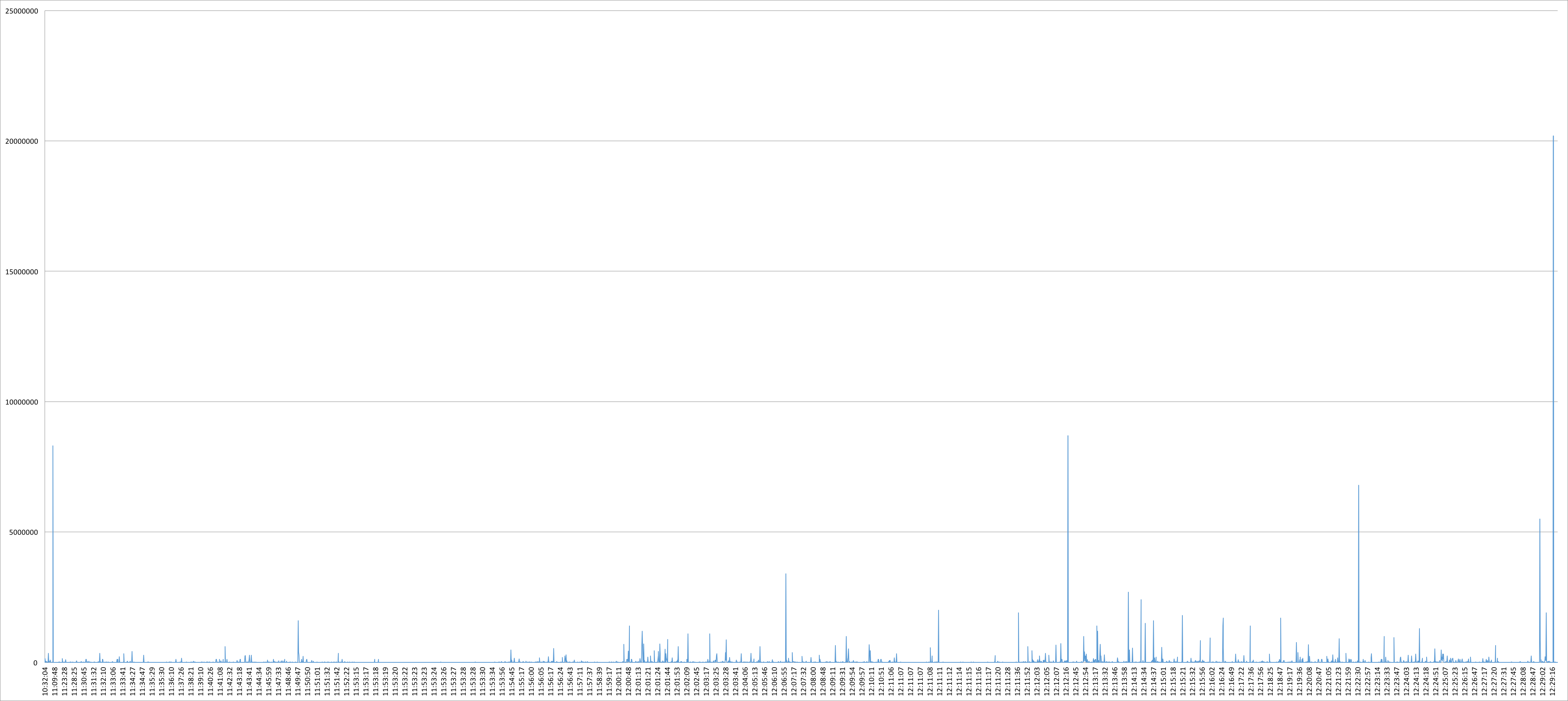
Далее для извлечения нужных данных, а именно объема трафика( Bytes) и времени (Date first seen), был написан код, листинг которого представлен на рисунке 5. На выходе данной программы имеем файл ***grep*** наполовину состоящий из величин объема трафика и на половину из значений времени, соответствующих объемам трафика. Также еще одним результатом выполнения данного кода является сумма всех объемов трафика (в байтах) (рисунок 6). Сравнивая данный результат (=176,81 МБ) я тем, который мы получили из статистики (рис.1, =176,7 МБ), можно отметить, что в статистике результаты посчитаны менее точно, однако погрешность, в переводе на денежный эквивалент согласно варианту задания, незначительная.

Получив необходимые для построения графика значения объема трафика и времени, перенесем их в таблицу excel и средствами MS-Excel построим требуемый график (рисунок 7).

**Выводы:** на основе проделанной лабораторной работы можно сделать вывод о том, что автоматизация процесса тарификации абонентов за услуги типа «Интернет» неизбежна, так же, как и за услуги типа «Телефония», поскольку количество активностей пользователей сети Интернет (манипулирование трафиком) значительно превышает само количество пользователей, что приводит к необходимости постоянной фильтрации и отбора данных, что сделать намного проще с помощью автоматизированных программных модулей. Таким образом непосредственно процесс тарификации не является сложным, поскольку не содержит в себе методов из сложных разделов математики, однако главной трудностью является именно количество обрабатываемой информации.

*Рис.5 «Листинг программы* ***graph-points.py****, выделяющей значения точек для построения графика»*

*Рис.6 «Суммарный объем трафика пользователя с IP-адресом 217.15.20.194 »*



*Рис.7 «График зависимости объема трафика от времени» (перевернут на 90о влево)*