Министерство образования и науки Российской Федерации

Федеральное государственное автономное образовательное учреждение высшего образования

**НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ ИТМО**



Факультет Безопасных Информационных Технологий

Управление мобильными устройствами

**Лабораторная работа №2**

**Выполнил:**

студент группы N3348

Колесников Н. Д.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Проверил:**

Федоров И. Р.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Санкт-Петербург

2020

**Цель работы**

В данной работе необходимо обработать трафик NetFlow v5 из файла nfcapd.202002251200. Для чего изначально требуется привести данный файл в читабельный вид (проще всего это сделать с помощью утилиты nfdump), после чего нужно сформировать собственный файл для тарификации любого формата, с которым удобно работать (в соответствии с вариантом работы), после чего необходимо построить график зависимости объема трафика от времени (любым удобным образом) и наконец требуется протарифицировать трафик в соответствии с вариантом задания.

**Средство реализации**

Для реализации мною был выбран язык программирования Python 3.x , т. к. я считаю что для выполнения данной задачи, он является наиболее подходящим вариантом (Так как присутствует большое количество модулей, используемых для работы с данными и их визуализации).

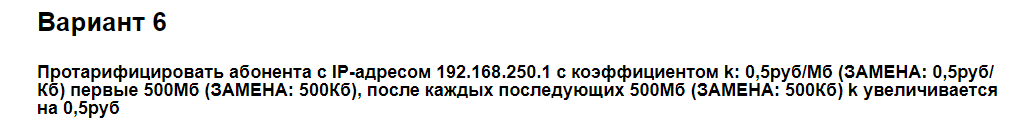
Для работы была выбрана среда Jupyter Notebook (также известный как IPython Botebook), которая крайне удобна в случае, если ты используешь Python (Так как весь код разделён на секции, что 1. упрощает процесс отладки программы 2. позволяет при необходимости выводить содержимое данных в любом месте программы).

**Исходный код**

<https://github.com/LordKzoth/MOBILE_MANAGEMENT/blob/master/LW_2/LW_2.py>

**Ход работы (Вариант 6)**

Задание:



Данные замены в задание были произведены согласно примечанию №2. Так как в последующем было выяснено, что общий трафик для данного IP приблизительно равен 6 Мб, что значительно меньше 500 Мб.

Ход работы:

Изначально с помощью команды «nfdump -r nfcapd.202002251200» был сформирован файл «nfcapd.txt».

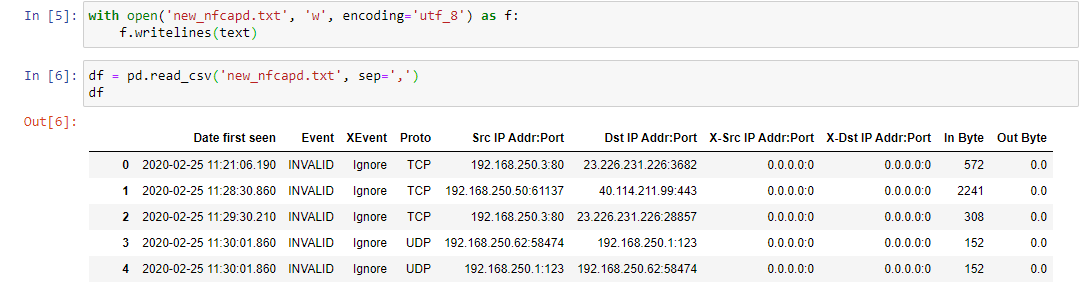
Следующим шагом этот файл необходимо было обработать таким образом, чтобы после этого его можно было бы прочитать с помощью модуля pandas.



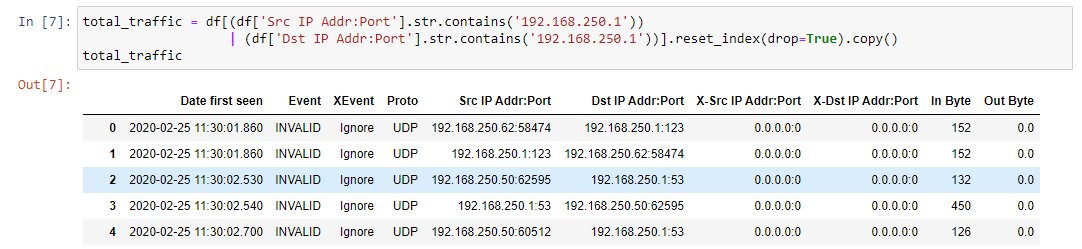
Описание данного блока:

1. Содержимое файла «nfcapd.txt» записывается в переменную text;
2. Далее это содержимое обрабатывается следующим образом:
   1. Очищается от символов «»;
   2. Добавляются дополнительные пробелы, там где располагаются одинарные (это необходимо для последующей очистки);
   3. Текст очищается от множественных пробелов, которые заменяются на двойные (это необходимо для того, чтобы можно было понять где стоит обыкновенный пробел, а где стоит пробел-разделитель);
   4. Двойные пробелы заменяются на запятые.

Следующим шагом отредактированный текст записывается в новый файл «new\_nfcapd.txt», который на следующем шаге обрабатывается с помощью метода pandas.read\_csv(). Результат записывается в переменную df.



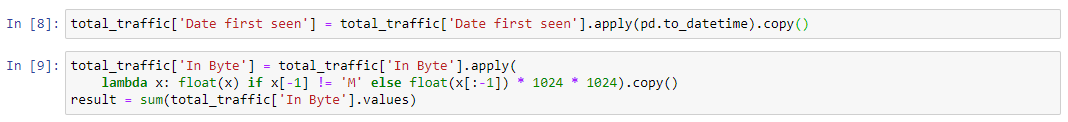
Следующим шагом из всех записей, были выбраны только те, в которых фигурирует IP=192.168.250.1



Таких записей оказалось 7064 штук.

На следующем шаге значения из столбика «Date first seen» были приведены к типу numpy.datetime, а также значения из столбика «In Byte», которые были записаны в виде «x.x М», были переведены в байты.

В завершение этого шага все значения из столбика «In Byte» были просуммированы.



Далее результат переводится в Кб.



Для отображения графика зависимости объёма данных от времени нам необходимо импортировать модуль matplotlib.



Для построения графика нам необходимо сгруппировать записи по времени (значения в столбике «In Byte» суммируются по группам).



Далее строится график (Данные были сгруппированы по миллисекундам).



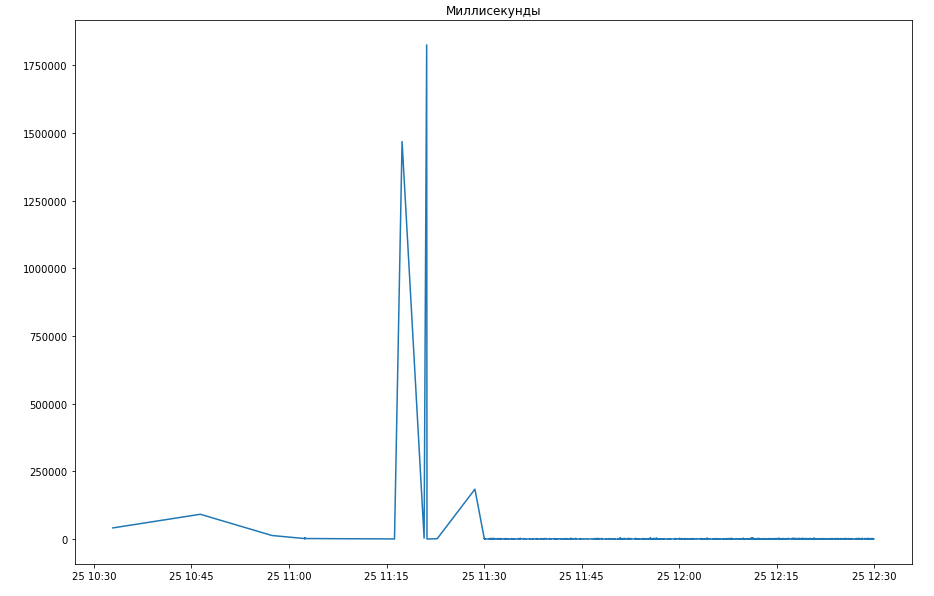


Рисунок 1 – График зависимости объёма трафика от времени (Миллисекунды)

Однако, на данном графике не сильно просматривается данная зависимость.

Следующим шагом было принято решение, сгруппировать данные по секундам и по минутам.

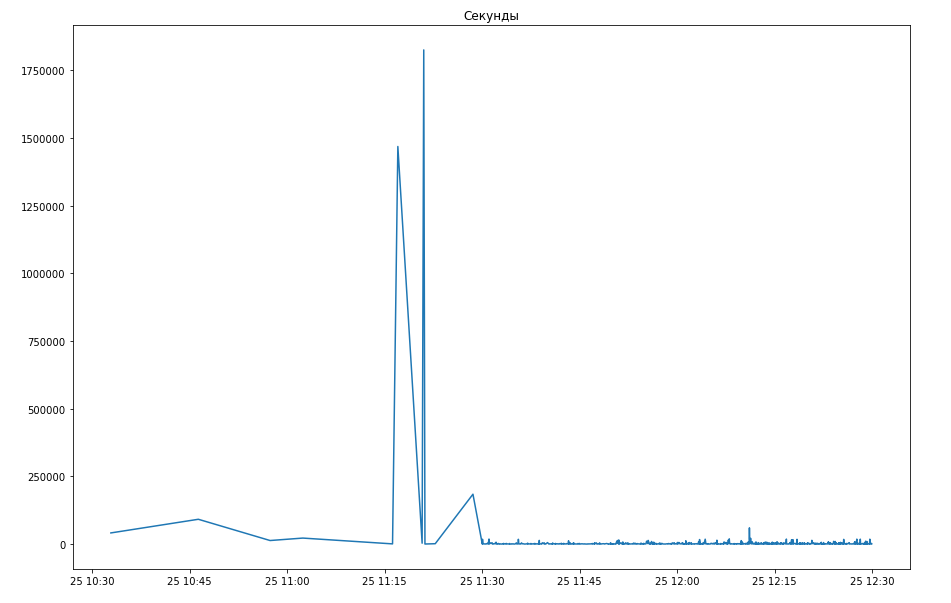


Рисунок 2 – График зависимости объёма трафика от времени (Секунды)

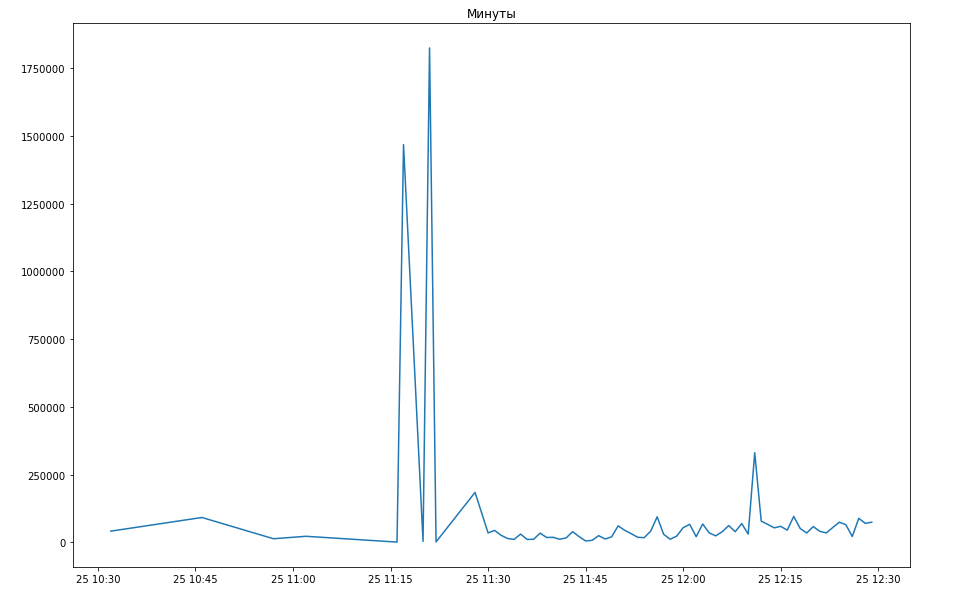
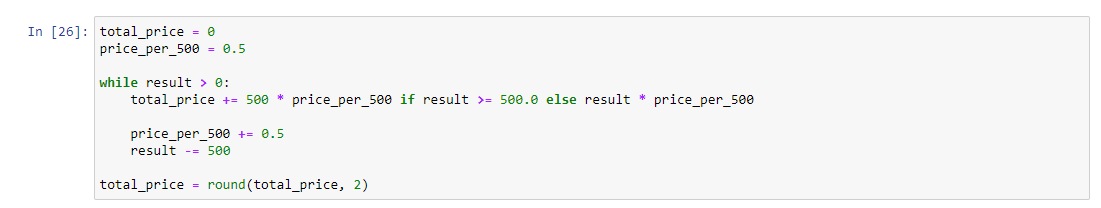


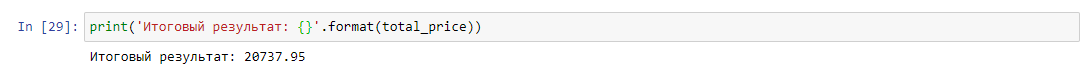
Рисунок 3 – График зависимости объёма трафика от времени (Минуты)

Наиболее четко зависимость просматривается на последнем графике, поэтому было принято решение использовать его в качестве основного.

На последнем шаге необходимо было произвести тарификацию согласно варианту.



**Итоговый результат**



Видим результат, округленный до сотых: **20737.95**

**Вывод:** в ходе выполнения данной лабораторной работы мною обследован предоставленный файл «nfcapd.202002251200», который был преобразован в файл «nfcapd.txt», который в свою очередь был преобразован в более удобный для анализа вид, который записан в итоговый файл «new\_nfcapd.txt», после чего из него были извлечены необходимые данные. После чего было реализовано простейшее правило тарификации для услуг типа «Интернет» по общему объёму трафика.