

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РФ

Федеральное государственное автономное
образовательное учреждение высшего образования
«Национальный исследовательский университет ИТМО»

ФАКУЛЬТЕТ БЕЗОПАСНОСТИ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ

Лабораторная работа №2
по дисциплине “Управление мобильными устройствами”
Вариант 11

Работу выполнила
студентка группы №3351
очного отделения
Лесина Валерия



Проверил:

Цель работы: реализация программного модуля для обработки, просмотра статистики и тарификации трафика NetFlow.

Задача:

Протарифицировать абонента с IP-адресом 17.248.150.51 с коэффициентом к: 0,5руб/Мб*

Ход работы:

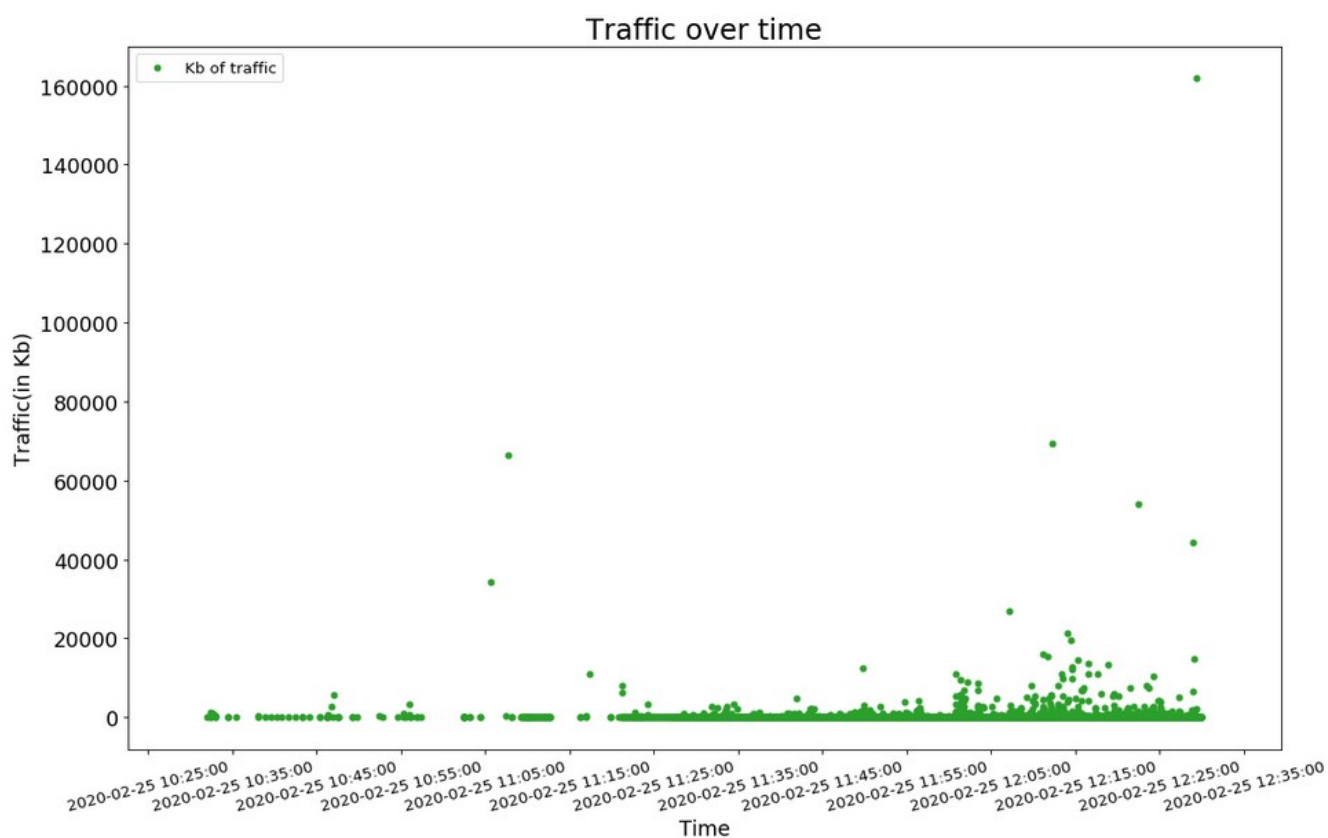
Исходя из тех же соображений, что и в предыдущей лабораторной работе, в качестве средств реализации данной задачи были выбраны язык программирования Python, библиотека pandas, а также библиотека matplotlib для построения графика зависимости объема трафика от времени.

Привести данный файл дампа трафика в читаемый вид можно командой `nfdump -r filename > dump.txt`. В дальнейшем напрямую с этим файлом работать не будем. Сформируем из дампа файл формата csv, с которым будет удобно работать библиотеке pandas, с помощью команды `nfdump -r nfcapd.202002251200 -o csv > dump.csv`. При попытке прочитать этот файл обнаруживается значительно большее количество столбцов, чем наблюдаемое в файле `dump.txt`. Обратившись к документации утилиты `nfdump` и сверив данные с “читабельным” выводом, приходим к выводу, что нам потребуются столбцы `sa,da,ibyt` и `ts`, обозначающие IP-адрес отправителя, IP-адрес получателя, количество переданных байт и время отправки данных соответственно (данные из последнего столбца необходимы для построения требуемого графика).

В ходе анализа данных по требуемому абоненту можно заметить, что общий объем трафика, в котором тот участвует, не превышает 1Мб, заявленного в первоначальном условии тарификации. В целях наглядности правило тарификации было изменено на 0.5руб/Кбит. Также, из условия задания неясно, о каких единицах измерения идет речь (мегабайт принято обозначать как МБ, а мегабит – как Мбит), поэтому расчеты были сделаны для единиц измерения Мбит(Кбит) –

исходя из существующих на рынке тарифных планов различных провайдеров.

Для построения графика зависимости общего объема трафика от времени, как уже было упомянуто ранее, была использована библиотека matplotlib. Было решено построить точечный график, являющийся классическим видом диаграммы, используемым для изучения корреляции двух переменных. В качестве используемых переменных были выбраны ts – время отправки данных, и ibyt – количество отправленных данных в байтах. На графике данные ibyt представлены в виде килобит.



Можно заметить, что начиная с 11:05, растет частота генерации трафика: данные отправляются чаще, а примерно с 11:45 увеличивается и их средний размер. Тем не менее, передача хостами большого объема данных достаточно редка – можно наблюдать лишь один случай передачи данных, чей размер превышает 16000 Кбит или 2 МБ.

Исходный код:

plot.py:

```
import pandas as pd
import matplotlib
import matplotlib.pyplot as plt
import matplotlib.dates as md
import datetime
import matplotlib.patches as patches

large = 22; med = 16; small = 12
params = {'axes.titlesize': large,
          'legend.fontsize': med,
          'figure.figsize': (16, 10),
          'axes.labelsize': med,
          'axes.titlesize': med,
          'xtick.labelsize': small,
          'ytick.labelsize': med,
          'figure.titlesize': large}
plt.rcParams.update(params)
coef = 0.008
df = pd.read_csv("dump.csv", low_memory=False)
x = [datetime.datetime.strptime(s, "%Y-%m-%d %H:%M:%S") for s in df["ts"][17449]]
y = [b*coef for b in df["ibyt"][17449]]
datenums=md.date2num(x)

plt.figure(figsize=(16, 10), dpi= 80, facecolor='w', edgecolor='k')
plt.xticks( rotation=14)
ax = plt.gca()
ax.set(xlabel='Time', ylabel='Traffic(in Kb)')
xfmt = md.DateFormatter('%Y-%m-%d %H:%M:%S')
ax.xaxis.set_major_formatter(xfmt)
ax.xaxis.set_major_locator(md.MinuteLocator(interval=10))
plt.scatter(datenums,y,s=20,c='tab:green',label='Kb of traffic')
plt.title("Traffic over time", fontsize=22)
plt.legend(fontsize=12)
#plt.show()
plt.savefig("stats.pdf",bbox_inches='tight')
```

lab2.py:

```
import pandas as pd
df = pd.read_csv("dump.csv", low_memory=False)
#ibyt ~ bytes field
#sa - source addr
#da - dest addr
```

```
l = len(df.index)
coef = 0.008 # bytes to kilobits
S = 0
k = 0.5
for i in range(0,l):
    if df["sa"][i] == "17.248.150.51" or df["da"][i] == "17.248.150.51" :
        S += df["ibyt"][i]
final=k*S*coef
print("Итоговый счет: {} рублей".format(final))
```

Ссылка на репозиторий github: <https://github.com/lesshuman/mobile/tree/master/lab2>

Выводы:

В данной работе были получены представления о протоколе NetFlow, методах его обработки с помощью утилиты nfdump, тарификации услуг типа “Интернет”. Были получены навыки работы с библиотекой matplotlib в ходе построения графика зависимости объема трафика от времени. Несмотря на то, что matplotlib позволяет строить разнообразные и наглядные графики, ее использование нельзя назвать интуитивным. Кроме того, отрисовка графиков на больших объемах данных требует заметного количества времени. Что касается библиотеки pandas, то она хорошо демонстрирует себя при работе со значительном объемом данных и позволяет произвести тарификацию без затруднений.

