

Нижегородский государственный технический университет им.

Р. Е. Алексеева

Институт радиоэлектроники и информационных технологий

Кафедра «Вычислительные системы и технологии»

Выпускная квалификационная работа

ПРОГРАММНАЯ СИСТЕМА АНАЛИЗА СЕТЕВОГО ТРАФИКА

Студент: Процкая Е.П. 15-В-1

Научный руководитель: к.т.н., доцент Гай В.Е.

Нижний Новгород 2019

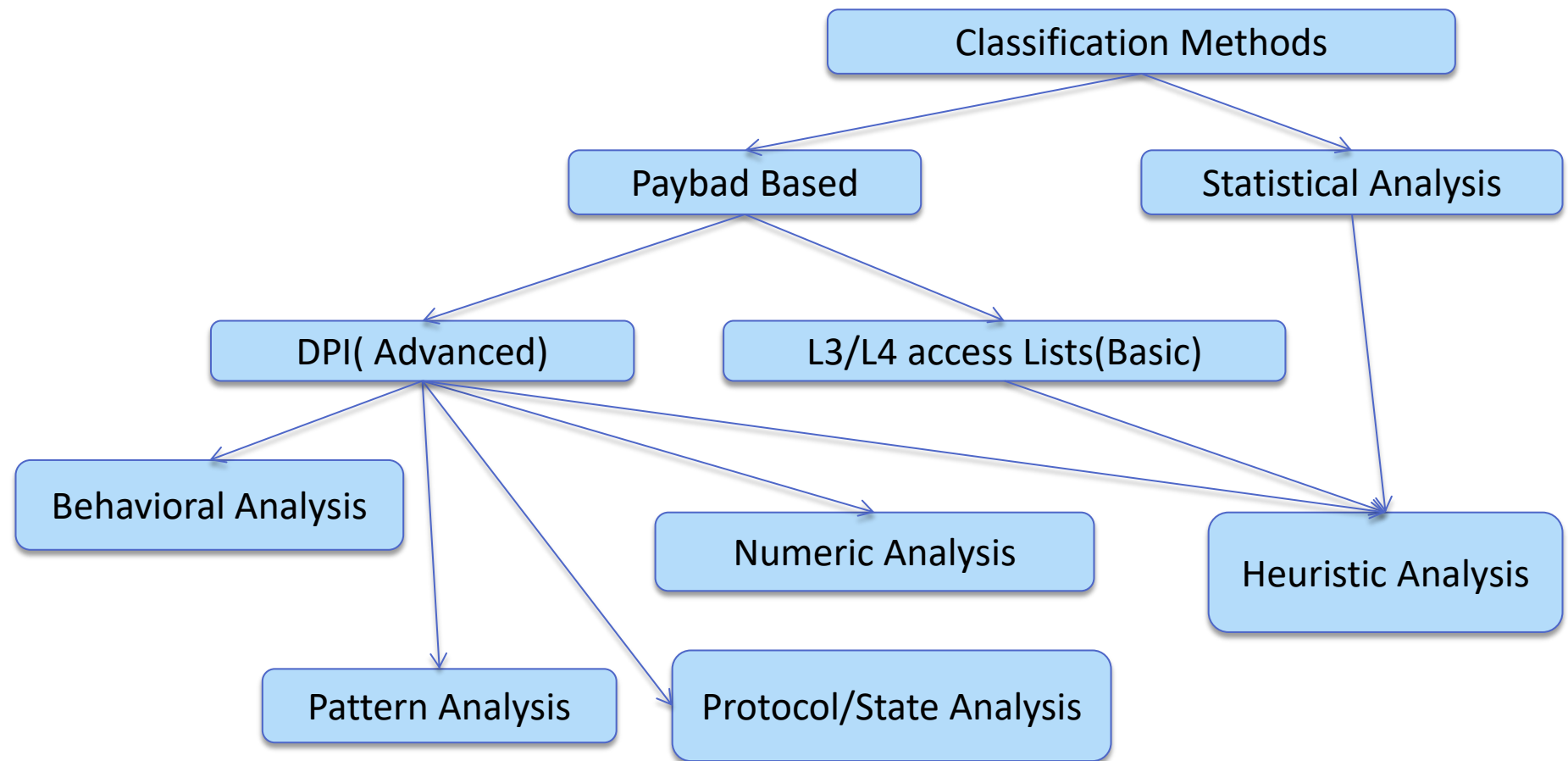
Цель и задачи исследования

❖ **Цель:** Автоматический анализ трафика.

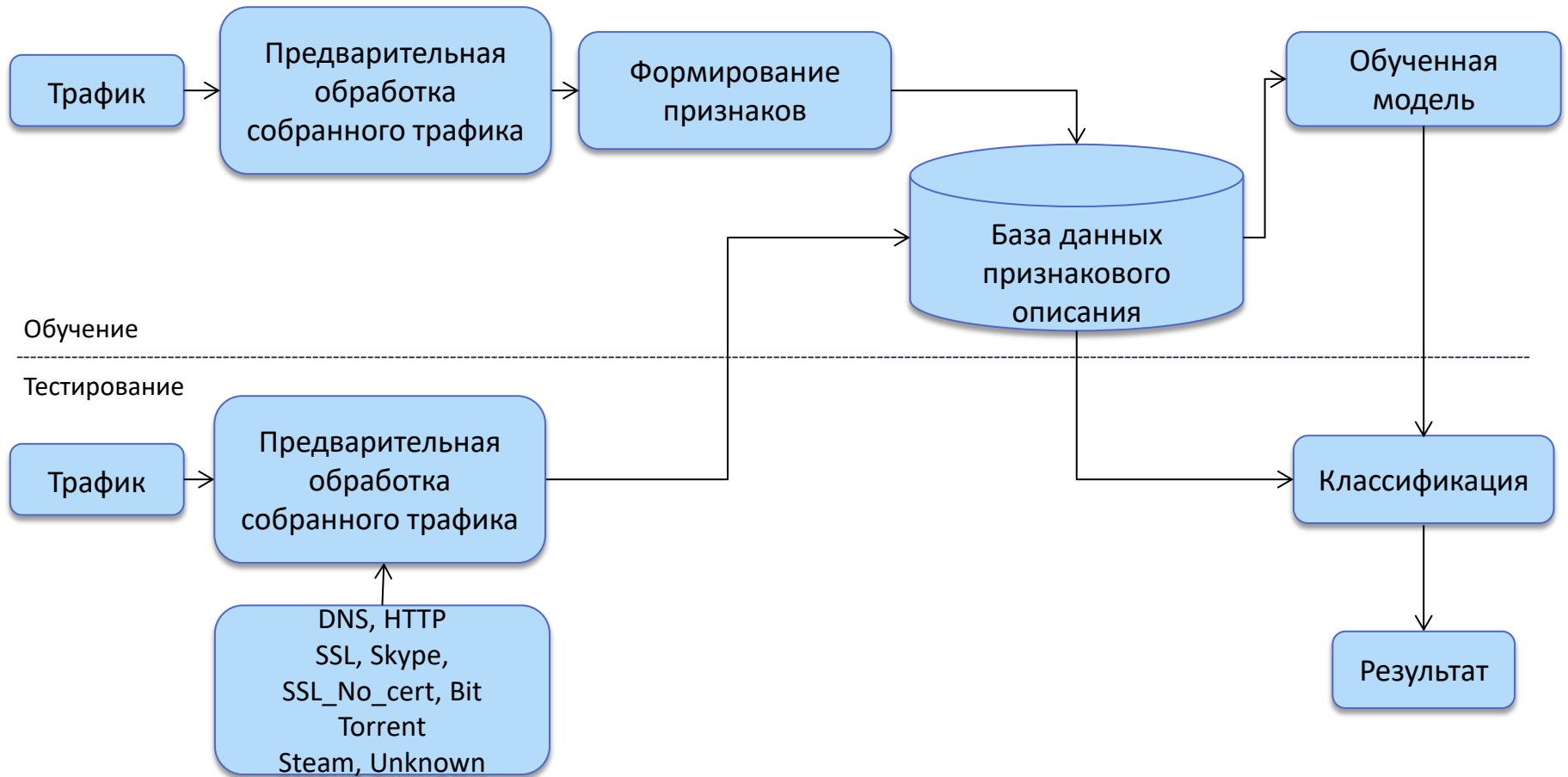
❖ **Задачи:**

1. Обзор методов классификации сетевого трафика;
2. Сбор трафика;
3. Тестирование алгоритмов машинного обучения в Azure ML, выбор подходящего алгоритма;
4. Разработка выбранного алгоритма классификации сетевого трафика ;
5. Проведение эксперимента для подтверждения корректности работы созданного алгоритма.

Подходы к решению задачи классификации



Информационная модель системы анализа сетевого трафика



Предварительная обработка исследуемого трафика

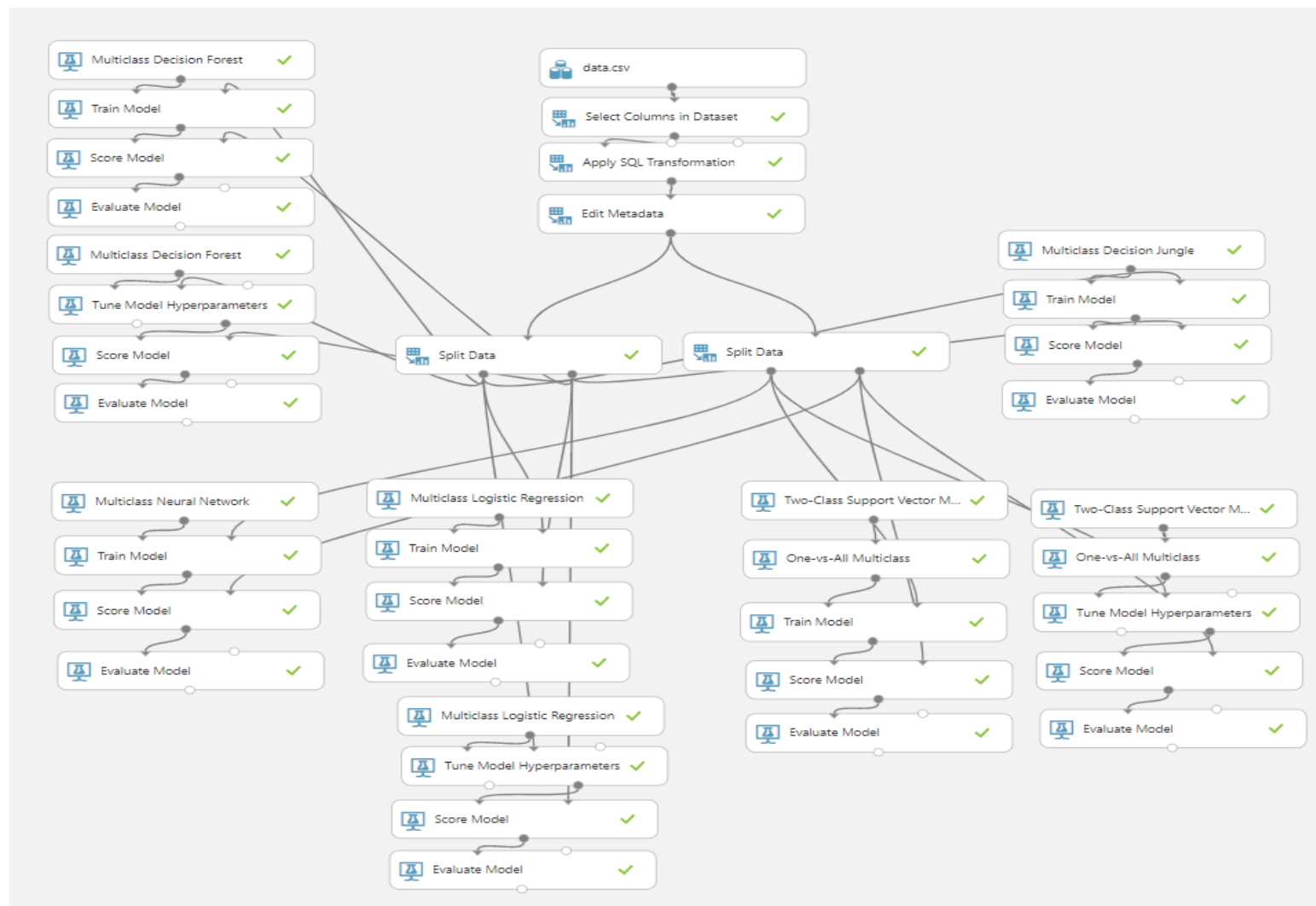
- ▶ Удалить из собранного набора данных пакеты относящиеся к редко используемым протоколам;
- ▶ Убрать протоколы не входящие в данный список: DNS, HTTP, SSL, Skype, Bit Torrent, Steam, Unknown;
- ▶ Объединить данные относящиеся к различным версиям одного протокола.

Формирование признаков

Используемые признаки:

- Последовательность размеров сегментов транспортного уровня (TCP или UDP), отправленных со стороны клиента;
- Последовательность размеров сегментов транспортного уровня, отправленных со стороны сервера;
- Последовательность размеров порций данных, отправленных со стороны клиента;
- Последовательность размеров порций данных, отправленных со стороны сервера.

Исследование в Microsoft Azure ML.

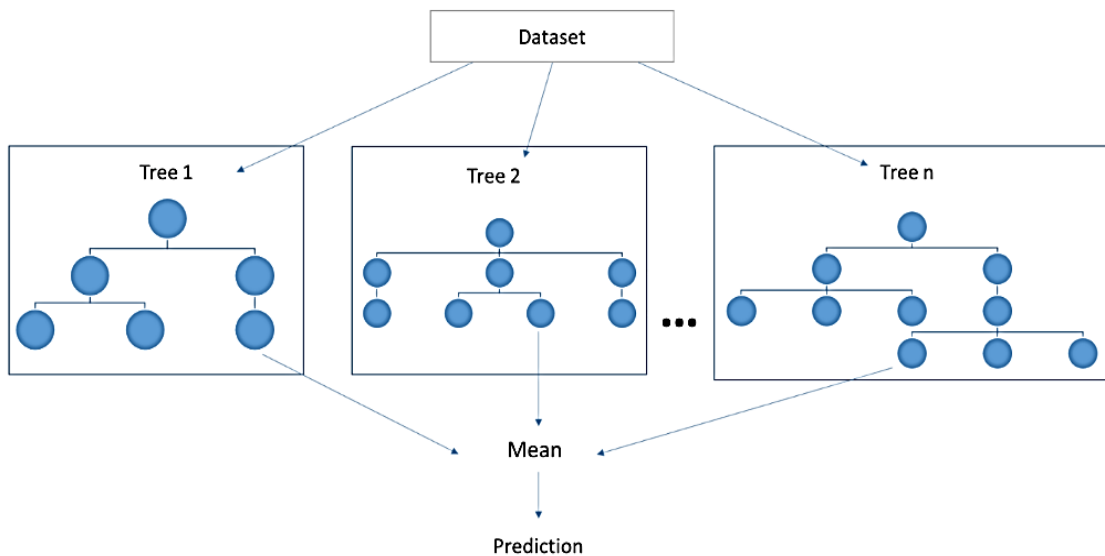


Результаты эксперимента

Классификатор/ значения	Точность при стандартных параметрах	Точность при лучших параметрах
Multiclass Decision Forest	0.962307	0.974064
Multiclass Decision Jungle	0.795524	0.951715
Multiclass Logistic Regression	0.988619	0.992792
Multiclass Neural Network	0.981034	-
One-vs-All Multiclass	0.961057	0.971275

Алгоритм «Random forest»

- ▶ Берутся обучающий наборы;
- ▶ Генерируется случайная выборка;
- ▶ Строится дерево принятия решений;
- ▶ Дерево строится до полного исчерпания выборки.



Параметры эксперимента:

Число деревьев - 32

Максимальная глубина дерева - 8

Принятие решения

- ▶ Признаковые описания классифицируют собранный трафик по выбранным протоколам;
- ▶ Исследуется таблица характеристик классифицированного трафика;
- ▶ Решение о точности классификации принимается на основе числа правильных классификаций трафика и числа ошибок.

Вычислительный эксперимент

- ▶ С помощью программы Wireshark произведен захват 10 гб трафика и отображены следующие виды DNS, BitTorrent, HTTP(S), SSL, Skype, Steam.
- ▶ Для выделения этих видов среди прочего трафика использована библиотека nDPI, осуществляющая глубокий анализ пакетов.
- ▶ Для запуска программный продукт необходим интерпретатор Python версии 2.7.x с установленными библиотеками PyQt4, dpkt, numpy, sklearn и pandas.

Тестирование программной системы

The image shows a terminal window and a file manager window. The terminal window displays test results for a software system, including precision, recall, f1-score, and support for various protocols, and a confusion matrix. The file manager window shows the directory structure of the project, including pcap files, CSV files, and model files.

Terminal Output:

```
server_efficiency 0.0064
client_packets 0.0062
server_packets 0.0057
bulk3 0.0055
byte_ratio 0.0045
server_packetsize_dev 0.0045
payload_ratio 0.0044
server_bulks 0.0025
```

	precision	recall	f1-score	support
BitTorrent	0.97	0.99	0.98	1156
DNS	1.00	0.99	1.00	364
HTTP	0.89	0.86	0.87	112
SSL	0.99	0.99	0.99	1126
Skype	0.50	0.07	0.12	14
Steam	1.00	0.75	0.86	4
Viber	0.00	0.00	0.00	10
avg / total	0.97	0.98	0.98	2786

	BitTorrent	DNS	HTTP	SSL	Skype	Steam	Viber
BitTorrent	1148	0	8	0	0	0	0
DNS	3	361	0	0	0	0	0
HTTP	3	0	96	13	0	0	0
SSL	3	0	3	1120	0	0	0
Skype	12	0	0	0	1	0	1
Steam	0	0	1	0	0	3	0
Viber	9	0	0	0	1	0	0

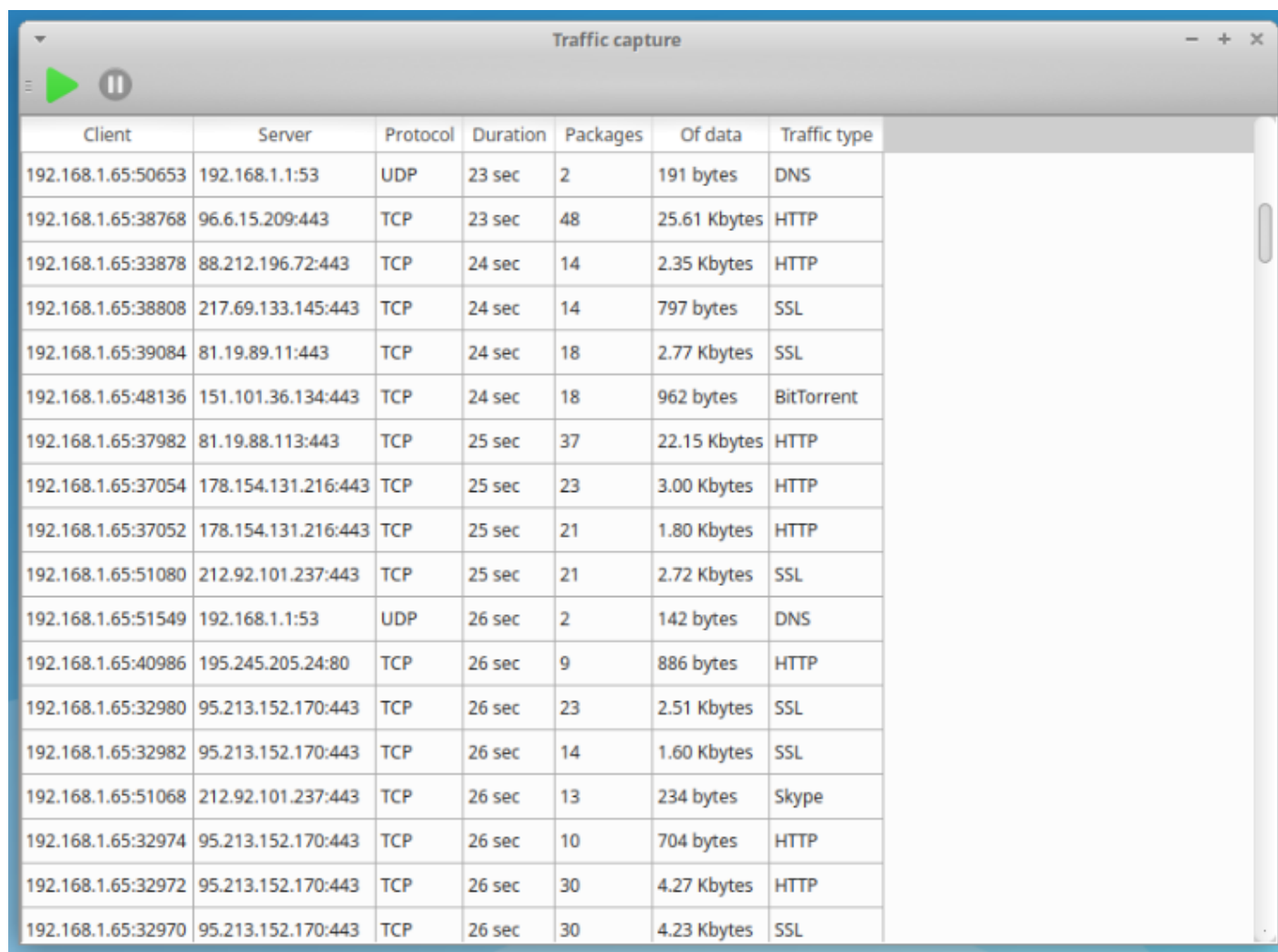
Module successfully written to file 'model-s1000.mdl'.
root@elizabeth-pc: /usr/TA/module_1_2#

File Manager: The file manager shows the directory structure of the project, including pcap files, CSV files, and model files.

- mdl
- pcaptocsv.py
- part0.pcap
- part1.pcap
- part2.pcap
- part3.pcap
- part4.pcap
- part5.pcap
- data-s10.csv
- data-s100.csv
- data-s1000.csv
- model.py
- model-s10.mdl
- model-s100.mdl
- model-s1000.mdl

12 объекта (21,4 ГБ), свободно 426,1 ГБ

Результат работы приложения



Client	Server	Protocol	Duration	Packages	Of data	Traffic type
192.168.1.65:50653	192.168.1.1:53	UDP	23 sec	2	191 bytes	DNS
192.168.1.65:38768	96.6.15.209:443	TCP	23 sec	48	25.61 Kbytes	HTTP
192.168.1.65:33878	88.212.196.72:443	TCP	24 sec	14	2.35 Kbytes	HTTP
192.168.1.65:38808	217.69.133.145:443	TCP	24 sec	14	797 bytes	SSL
192.168.1.65:39084	81.19.89.11:443	TCP	24 sec	18	2.77 Kbytes	SSL
192.168.1.65:48136	151.101.36.134:443	TCP	24 sec	18	962 bytes	BitTorrent
192.168.1.65:37982	81.19.88.113:443	TCP	25 sec	37	22.15 Kbytes	HTTP
192.168.1.65:37054	178.154.131.216:443	TCP	25 sec	23	3.00 Kbytes	HTTP
192.168.1.65:37052	178.154.131.216:443	TCP	25 sec	21	1.80 Kbytes	HTTP
192.168.1.65:51080	212.92.101.237:443	TCP	25 sec	21	2.72 Kbytes	SSL
192.168.1.65:51549	192.168.1.1:53	UDP	26 sec	2	142 bytes	DNS
192.168.1.65:40986	195.245.205.24:80	TCP	26 sec	9	886 bytes	HTTP
192.168.1.65:32980	95.213.152.170:443	TCP	26 sec	23	2.51 Kbytes	SSL
192.168.1.65:32982	95.213.152.170:443	TCP	26 sec	14	1.60 Kbytes	SSL
192.168.1.65:51068	212.92.101.237:443	TCP	26 sec	13	234 bytes	Skype
192.168.1.65:32974	95.213.152.170:443	TCP	26 sec	10	704 bytes	HTTP
192.168.1.65:32972	95.213.152.170:443	TCP	26 sec	30	4.27 Kbytes	HTTP
192.168.1.65:32970	95.213.152.170:443	TCP	26 sec	30	4.23 Kbytes	SSL

Публикации

- ▶ В.Е. Гай, П.Е. Процкая Программная система анализа сетевого трафика// Труды XXV международной конференции «Информационные системы и технологии» ИСТ-2019, 19 апреля 2019 г., С. 876-881

Спасибо за внимание !

