WILLIAM & MARY

CHARTERED 1693

Classification of Network Traffic by Website Function

Francis Adams, Saad Khan, Nicolette Glut, Elena Ceravalo, Joey Horner

Introduction

- Monitor and analyze network traffic
- Using features extracted from network traffic data to build machine learning models
- Use models to classify traffic to help understand user behavior, without tracking IP addresses
- Possible Application: blacklisting and whitelisting websites

Data collection

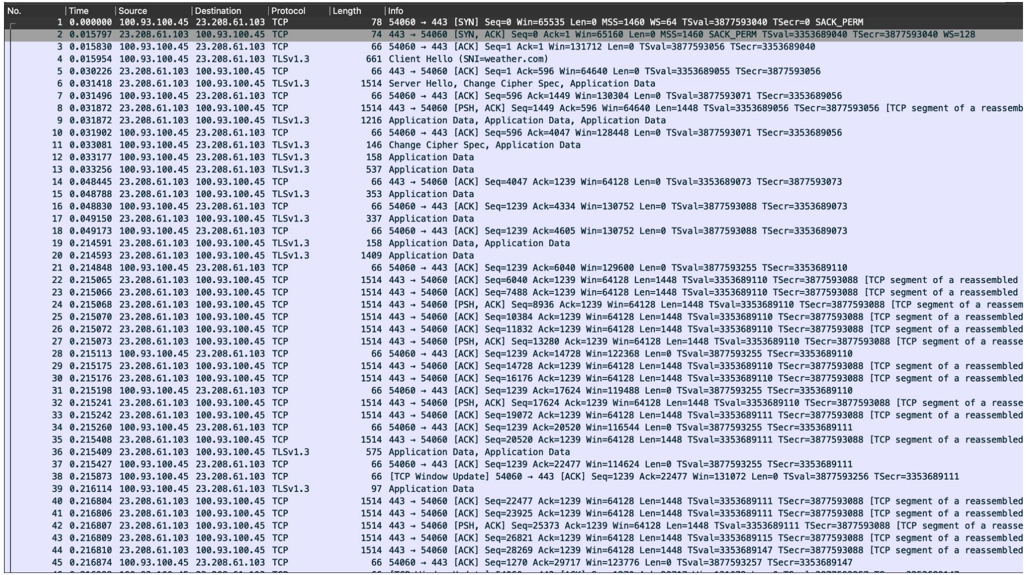
Extracted features: Protocol, Length, Info

Features:	**Details**
Protocol	Protocol used to communicate, i.e. TCP, UDP, TLSv1.3, etc.
Length	Size of a packet
Info	More details about communication of packet, i.e. ACK, TCP Duplicate ACK, Application Data, TCP Fast Retransmit, etc.

Dataset size: 38532 samples

Category	**Dataset size**
Shopping	10008
Social Media	9999
Streaming	9421
Banking	9104

Data collection



```
No.        Time     Source        Destination   Protocol     Length Info
         1 0.000000 100.93.100.45 23.208.61.103 TCP            78 54060 → 443 [SYN] Seq=0 Win=65535 Len=0 MSS=1460 WS=64 TSval=3877593040 TSecr=0 SACK_PERM
         2 0.015797 23.208.61.103 100.93.100.45 TCP            74 443 → 54060 [SYN, ACK] Seq=0 Ack=1 Win=65160 Len=0 MSS=1460 SACK_PERM TSval=3353689040 TSecr=3877593040 WS=128
         3 0.015830 100.93.100.45 23.208.61.103 TCP            66 54060 → 443 [ACK] Seq=1 Ack=1 Win=131712 Len=0 TSval=3877593056 TSecr=3353689040
         4 0.015954 100.93.100.45 23.208.61.103 TLSv1.3       661 Client Hello (SNI=weather.com)
         5 0.030226 23.208.61.103 100.93.100.45 TCP            66 443 → 54060 [ACK] Seq=1 Ack=596 Win=64640 Len=0 TSval=3353689055 TSecr=3877593056
         6 0.031418 23.208.61.103 100.93.100.45 TLSv1.3      1514 Server Hello, Change Cipher Spec, Application Data
         7 0.031496 100.93.100.45 23.208.61.103 TCP            66 54060 → 443 [ACK] Seq=596 Ack=1449 Win=130304 Len=0 TSval=3877593071 TSecr=3353689056
         8 0.031872 23.208.61.103 100.93.100.45 TCP          1514 443 → 54060 [PSH, ACK] Seq=1449 Ack=596 Win=64640 Len=1448 TSval=3353689056 TSecr=3877593056 [TCP segment of a reassemb
         9 0.031872 23.208.61.103 100.93.100.45 TLSv1.3      1216 Application Data, Application Data, Application Data
        10 0.031902 100.93.100.45 23.208.61.103 TCP            66 54060 → 443 [ACK] Seq=596 Ack=4047 Win=128448 Len=0 TSval=3877593071 TSecr=3353689056
        11 0.033081 100.93.100.45 23.208.61.103 TLSv1.3       146 Change Cipher Spec, Application Data
        12 0.033177 100.93.100.45 23.208.61.103 TLSv1.3       158 Application Data
        13 0.033256 100.93.100.45 23.208.61.103 TLSv1.3       537 Application Data
        14 0.048445 23.208.61.103 100.93.100.45 TCP            66 443 → 54060 [ACK] Seq=4047 Ack=1239 Win=64128 Len=0 TSval=3353689073 TSecr=3877593073
        15 0.048788 23.208.61.103 100.93.100.45 TLSv1.3       353 Application Data
        16 0.048830 100.93.100.45 23.208.61.103 TCP            66 54060 → 443 [ACK] Seq=1239 Ack=4334 Win=130752 Len=0 TSval=3877593088 TSecr=3353689073
        17 0.049150 23.208.61.103 100.93.100.45 TLSv1.3       337 Application Data
        18 0.049173 100.93.100.45 23.208.61.103 TCP            66 54060 → 443 [ACK] Seq=1239 Ack=4605 Win=130752 Len=0 TSval=3877593088 TSecr=3353689073
        19 0.214591 23.208.61.103 100.93.100.45 TLSv1.3       158 Application Data, Application Data
        20 0.214593 23.208.61.103 100.93.100.45 TLSv1.3      1409 Application Data
        21 0.214848 100.93.100.45 23.208.61.103 TCP            66 54060 → 443 [ACK] Seq=1239 Ack=6040 Win=129600 Len=0 TSval=3877593255 TSecr=3353689110
        22 0.215065 23.208.61.103 100.93.100.45 TCP          1514 443 → 54060 [ACK] Seq=6040 Ack=1239 Win=64128 Len=1448 TSval=3353689110 TSecr=3877593088 [TCP segment of a reassembled
        23 0.215066 23.208.61.103 100.93.100.45 TCP          1514 443 → 54060 [ACK] Seq=7488 Ack=1239 Win=64128 Len=1448 TSval=3353689110 TSecr=3877593088 [TCP segment of a reassembled
        24 0.215068 23.208.61.103 100.93.100.45 TCP          1514 443 → 54060 [PSH, ACK] Seq=8936 Ack=1239 Win=64128 Len=1448 TSval=3353689110 TSecr=3877593088 [TCP segment of a reassem
        25 0.215070 23.208.61.103 100.93.100.45 TCP          1514 443 → 54060 [ACK] Seq=10384 Ack=1239 Win=64128 Len=1448 TSval=3353689110 TSecr=3877593088 [TCP segment of a reassembled
        26 0.215072 23.208.61.103 100.93.100.45 TCP          1514 443 → 54060 [ACK] Seq=11832 Ack=1239 Win=64128 Len=1448 TSval=3353689110 TSecr=3877593088 [TCP segment of a reassembled
        27 0.215073 23.208.61.103 100.93.100.45 TCP          1514 443 → 54060 [PSH, ACK] Seq=13280 Ack=1239 Win=64128 Len=1448 TSval=3353689110 TSecr=3877593088 [TCP segment of a reasse
        28 0.215113 100.93.100.45 23.208.61.103 TCP            66 54060 → 443 [ACK] Seq=1239 Ack=14728 Win=122368 Len=0 TSval=3877593255 TSecr=3353689110
        29 0.215175 23.208.61.103 100.93.100.45 TCP          1514 443 → 54060 [ACK] Seq=14728 Ack=1239 Win=64128 Len=1448 TSval=3353689110 TSecr=3877593088 [TCP segment of a reassembled
        30 0.215176 23.208.61.103 100.93.100.45 TCP          1514 443 → 54060 [ACK] Seq=16176 Ack=1239 Win=64128 Len=1448 TSval=3353689110 TSecr=3877593088 [TCP segment of a reassembled
        31 0.215198 100.93.100.45 23.208.61.103 TCP            66 54060 → 443 [ACK] Seq=1239 Ack=17624 Win=119488 Len=0 TSval=3877593255 TSecr=3353689110
        32 0.215241 23.208.61.103 100.93.100.45 TCP          1514 443 → 54060 [PSH, ACK] Seq=17624 Ack=1239 Win=64128 Len=1448 TSval=3353689110 TSecr=3877593088 [TCP segment of a reasse
        33 0.215242 23.208.61.103 100.93.100.45 TCP          1514 443 → 54060 [ACK] Seq=19072 Ack=1239 Win=64128 Len=1448 TSval=3353689111 TSecr=3877593088 [TCP segment of a reassembled
        34 0.215260 100.93.100.45 23.208.61.103 TCP            66 54060 → 443 [ACK] Seq=1239 Ack=20520 Win=116544 Len=0 TSval=3877593255 TSecr=3353689111
        35 0.215408 23.208.61.103 100.93.100.45 TCP          1514 443 → 54060 [ACK] Seq=20520 Ack=1239 Win=64128 Len=1448 TSval=3353689111 TSecr=3877593088 [TCP segment of a reassembled
        36 0.215409 23.208.61.103 100.93.100.45 TLSv1.3       575 Application Data, Application Data
        37 0.215427 100.93.100.45 23.208.61.103 TCP            66 54060 → 443 [ACK] Seq=1239 Ack=22477 Win=114624 Len=0 TSval=3877593255 TSecr=3353689111
        38 0.215873 100.93.100.45 23.208.61.103 TCP            66 [TCP Window Update] 54060 → 443 [ACK] Seq=1239 Ack=22477 Win=131072 Len=0 TSval=3877593256 TSecr=3353689111
        39 0.216114 100.93.100.45 23.208.61.103 TLSv1.3        97 Application Data
        40 0.216804 23.208.61.103 100.93.100.45 TCP          1514 443 → 54060 [ACK] Seq=22477 Ack=1239 Win=64128 Len=1448 TSval=3353689111 TSecr=3877593088 [TCP segment of a reassembled
        41 0.216806 23.208.61.103 100.93.100.45 TCP          1514 443 → 54060 [ACK] Seq=23925 Ack=1239 Win=64128 Len=1448 TSval=3353689111 TSecr=3877593088 [TCP segment of a reassembled
        42 0.216807 23.208.61.103 100.93.100.45 TCP          1514 443 → 54060 [PSH, ACK] Seq=25373 Ack=1239 Win=64128 Len=1448 TSval=3353689111 TSecr=3877593088 [TCP segment of a reasse
        43 0.216809 23.208.61.103 100.93.100.45 TCP          1514 443 → 54060 [ACK] Seq=26821 Ack=1239 Win=64128 Len=1448 TSval=3353689115 TSecr=3877593088 [TCP segment of a reassembled
        44 0.216810 23.208.61.103 100.93.100.45 TCP          1514 443 → 54060 [ACK] Seq=28269 Ack=1239 Win=64128 Len=1448 TSval=3353689147 TSecr=3877593088 [TCP segment of a reassembled
        45 0.216874 100.93.100.45 23.208.61.103 TCP            66 54060 → 443 [ACK] Seq=1270 Ack=29717 Win=123776 Len=0 TSval=3877593257 TSecr=3353689147
```

Shopping (0)

Etsy

Uber

Social Media (1)

Streaming (2)

Banking (3)

TRUIST

WELLS FARGO

No.	Time	Source	Destination	Protocol	Length	Info	label
1	0.0	192.168.1.31	52.3.144.142	TLSv1.2	118	Application Data	0
2	2.8E-05	192.168.1.31	52.3.144.142	TLSv1.2	93	Application Data	0
3	0.016127	52.3.144.142	192.168.1.31	TCP	60	443 > 56483 [ACK] Seq=1 Ack=104 Win=68 Len=0	0
4	0.016873	52.3.144.142	192.168.1.31	TLSv1.2	93	Application Data	0
5	0.059521	192.168.1.31	52.3.144.142	TCP	54	56483 > 443 [ACK] Seq=104 Ack=40 Win=1025 Len=0	0

Model Architecture

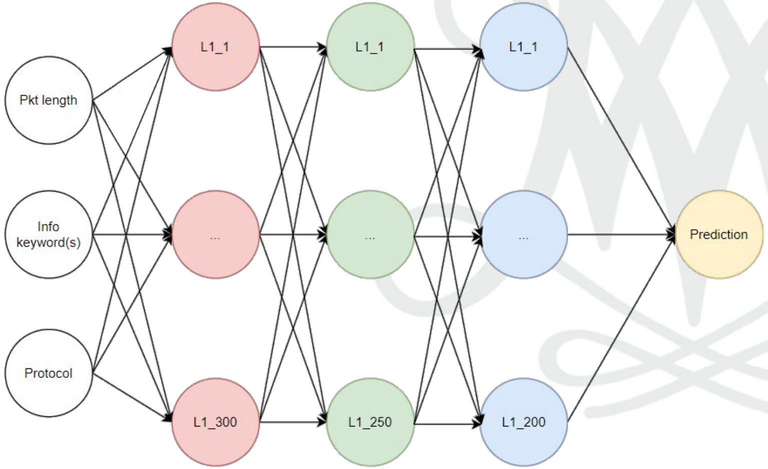
Decision Tree:

- Split classifications by threshold for feature at each parent node, final leaf node reached represents prediction class

Multilayer Perceptron (MLP):

- Model has inner nodes between original input and output associated with weights which are influenced by error in output at a node. Node weights are received by each node in the next layer

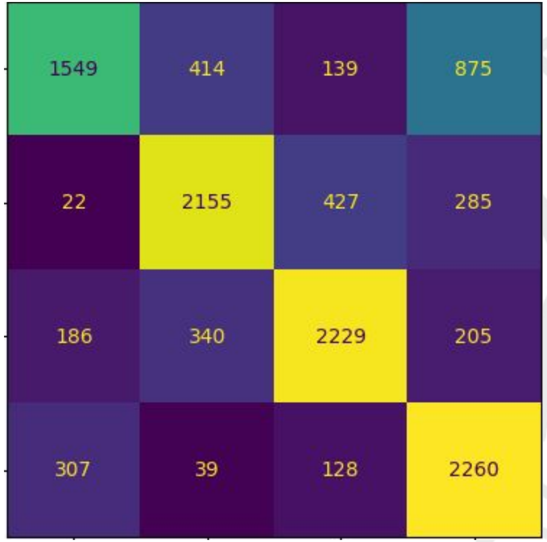
K Nearest Neighbors: Make classifications based on classes of K neighbors closest to data point. Closer neighbors have more influence



Evaluation & Results

	Decision Tree	**KNN**	**MLP**
Accuracy	0.709	0.700	0.537
F1 Score	0.705	0.701	0.510

Decision Tree Confusion Matrix



1549	414	139	875
22	2155	427	285
186	340	2229	205
307	39	128	2260