whalebone

Neural Networks and Challenges in Detection of Malicious DNS Traffic and DGA malware

Robert Šefr, CTO @ Whalebone
@robcza

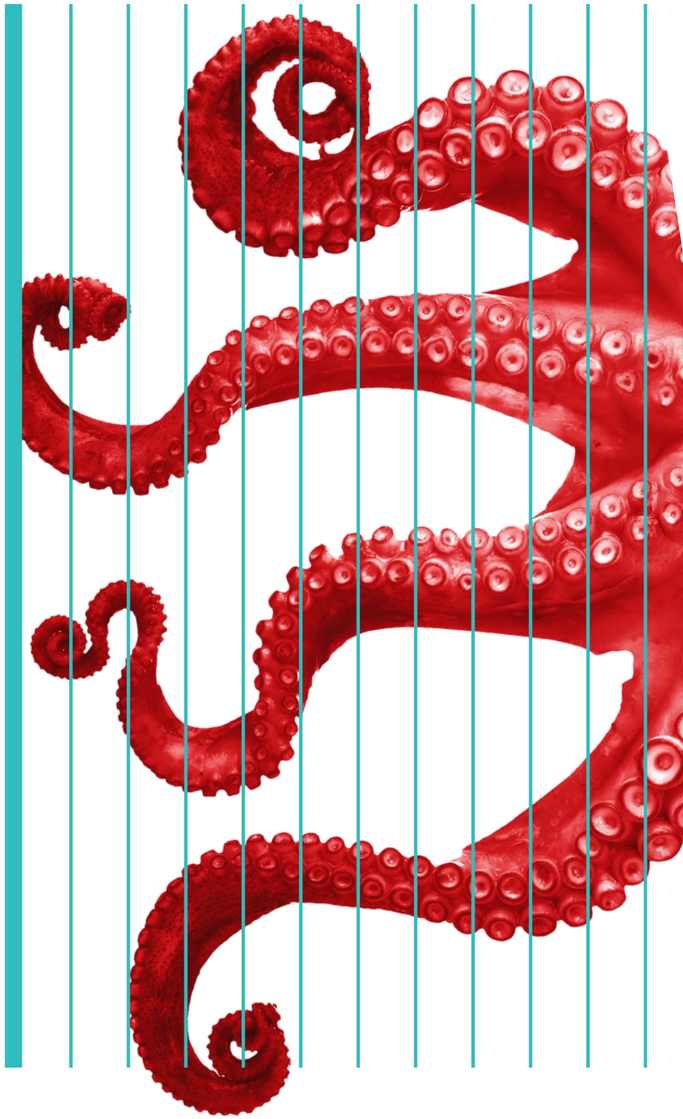
Domain Generation Algorithm (DGA) detection

- Detect sequences of malware generated domains among regular traffic
- Single IP can be a particular device or a whole large network
- DGAs can generate from several requests par day to hundreds of thousands

- **Ultimate goal – detect infected clients without prior knowledge of the particular DGA**

eygyww.com
bgraqmfuq.org
bcvfukmgemu.info
rgjzxqptp.net
gotmgqlkjkr.info
ttqeijmhwxvn.info
zotkihl.org
uoeigsoa.org
lalkfmb.org
rgejjlb.org
jgvcpgk.info
ztbljj.net
hsnuhtbgnia.net
fbtsjfew.net
msbkdwcer.net
fyxpvajh.info

Per domain DGA detection



Whalebone sponsored research

- **Carlos Catania, Sebastian García and Pablo Torres**
- Czech Technical University, Prague
- Universidad de Mendoza, Argentina

- 1,003,161 clean domains
- 1,915,335 DGA domains
- 51 malware families

Family	Scheme	Freq.	Family	Scheme	Freq.	Family	Scheme	Freq.
bamital	(A)	904	cryptolocker	(A)	112,809	padcrypt	(A)	1,920
p2p	(A)	4,000	proslikefan	(A)	100	murofet	(A)	49,199
bedep	(A)	706	dircrypt	(A)	570	necurs	(A)	81,920
post	(A)	220,000	dyre	(A)	26,993	newgoz	(A)	1,666
chinad	(A)	256	fobber	(A)	600	nymaim	(A)	20,225
conficker	(A)	99,996	gameover	(A)	12,000	pushdo	(A)	94,278
corebot	(A)	840	geodo	(A)	1,920	pykspa	(A)	25,727
goz	(A)	1,667	hesperbot	(A)	192	qadars	(A)	1,600
kraken	(A)	9,660	locky	(A)	9,028	qakbot	(A)	60,000
ramdo	(A)	102,000	ramnit	(A)	91,978	ranbyus	(A)	23,167
rovnix	(A)	53,632	shiotob	(A)	12,521	symmi	(A)	4,448
shifu	(A)	2,554	virut	(A)	11,994	sisron	(A)	60
zeus	(A)	1,000	vawtrak	(A)	300	simda	(A)	28,339
tinba	(A)	193,912	tempedreve	(A)	225	pykspa$_v$1	(A)	18
pykspav2F	(A)	800	pykspav2R	(A)	200	banjori	(W)	439218
suppobox	(W)	8185	matsnu	(W)	100127	volatile	(W)	996
beebone	(W)	210	cryptowall	(W)	94	madmax	(A)	2

https://link.springer.com/chapter/10.1007/978-3-030-20787-8_23

Neural Network Architecture

1. **Embedding layer**
   - Vectorize the input
2. **1D Convolutional layer**
   - Extract the features (4-grams)
3. **Multilayer Perceptron Network**
   - Calculate the probability and provide result

- The result is simple: “DGA” or “normal”

Layer (type)	Activation Function
input (Input Layer)	-
embedding (Embedding)	-
conv1d (1D Convolutional)	relu
dense_1 (Dense)	relu
dense_2 (Dense)	sigmoid

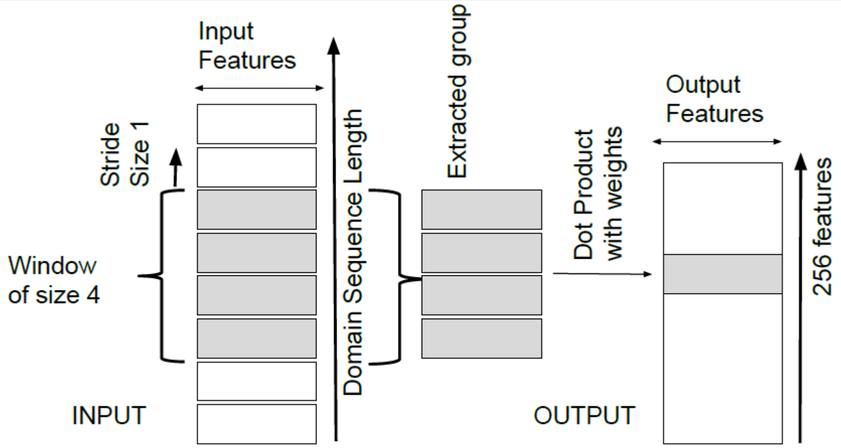


Fig. 1. Feature extraction process of the 1D Convolutional layer

Results

Original research

- True Positive rate: 97%
- False positive rate: 0,7%

Our current NN has significantly better results (more effort into manual labeling)

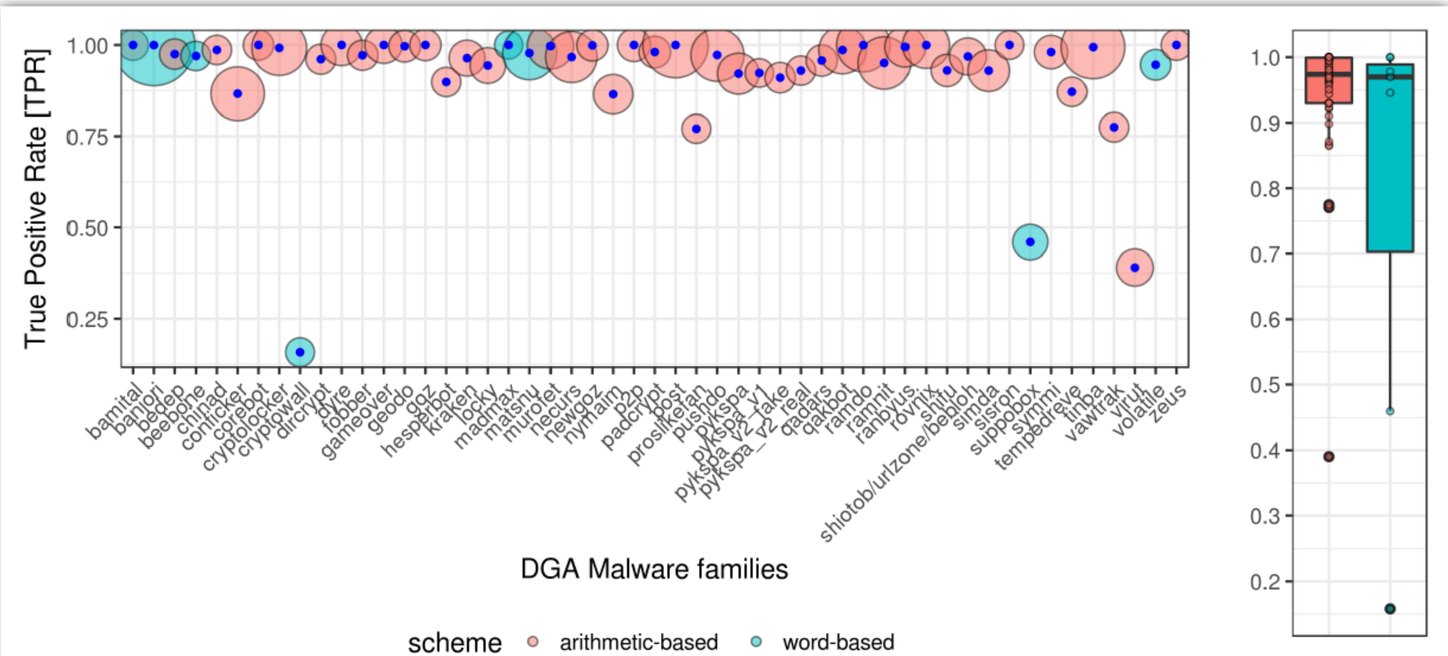
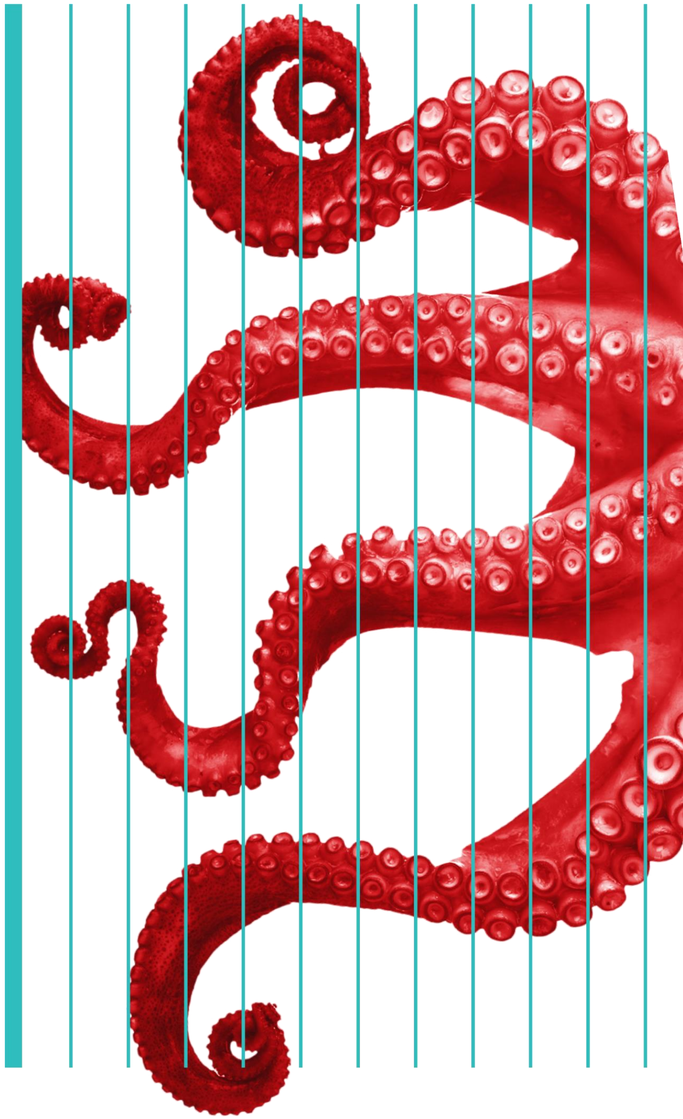


Fig. 2. True Positive Rate per DGA malware family for 1D-CNN

Infected client detection



The issue

- Through the NN we know the individual domain could be DGA
- It is impossible to have reasonably low False Positive rate on single domain - some domains were born to be FPs:

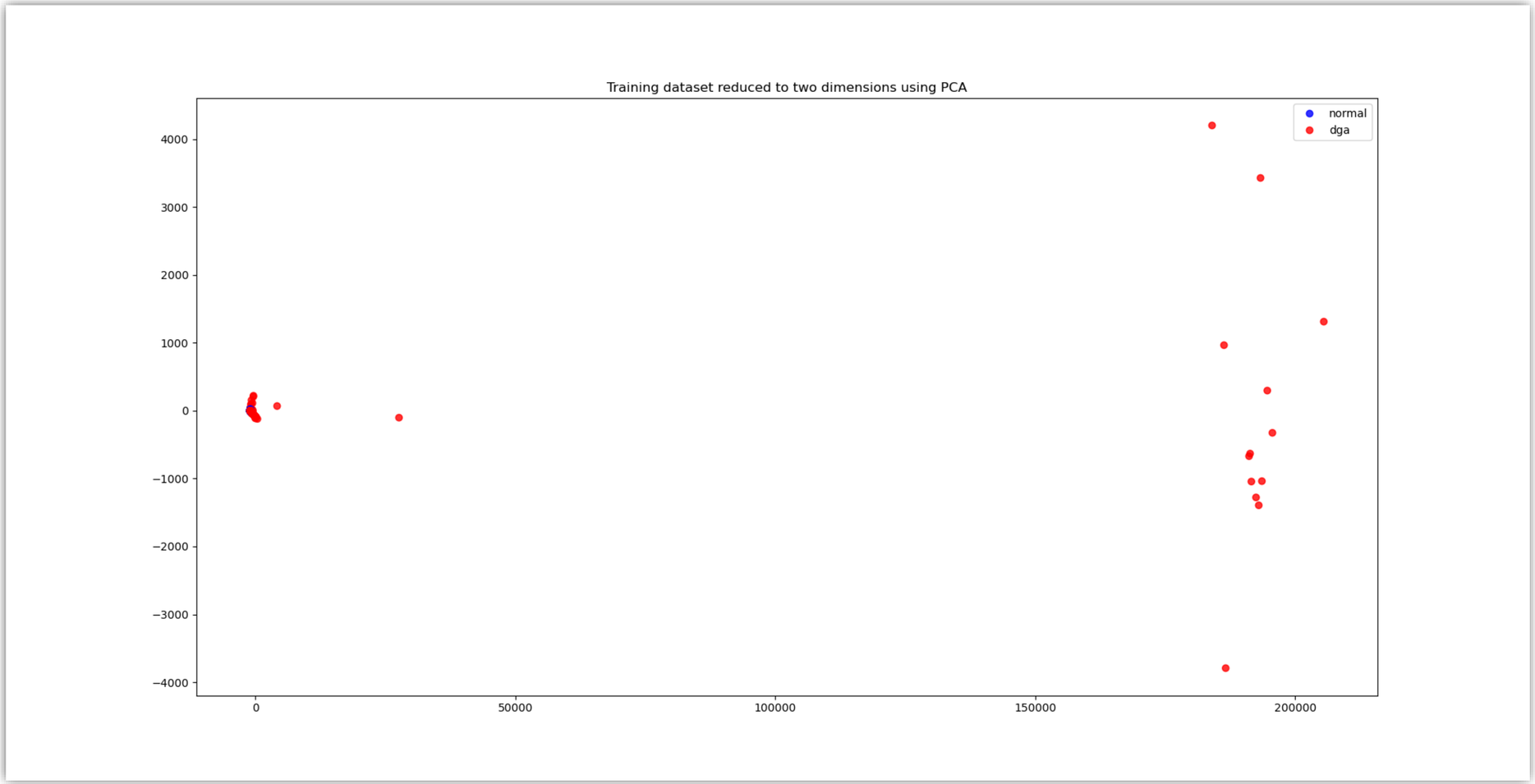
csvnmnm.cz
vospaspsm.cz
zbkjmkcr.cz
zusjjrrozmitalptr.cz

- At the end we don’t care about the domain, but about the client. **Is it infected?**

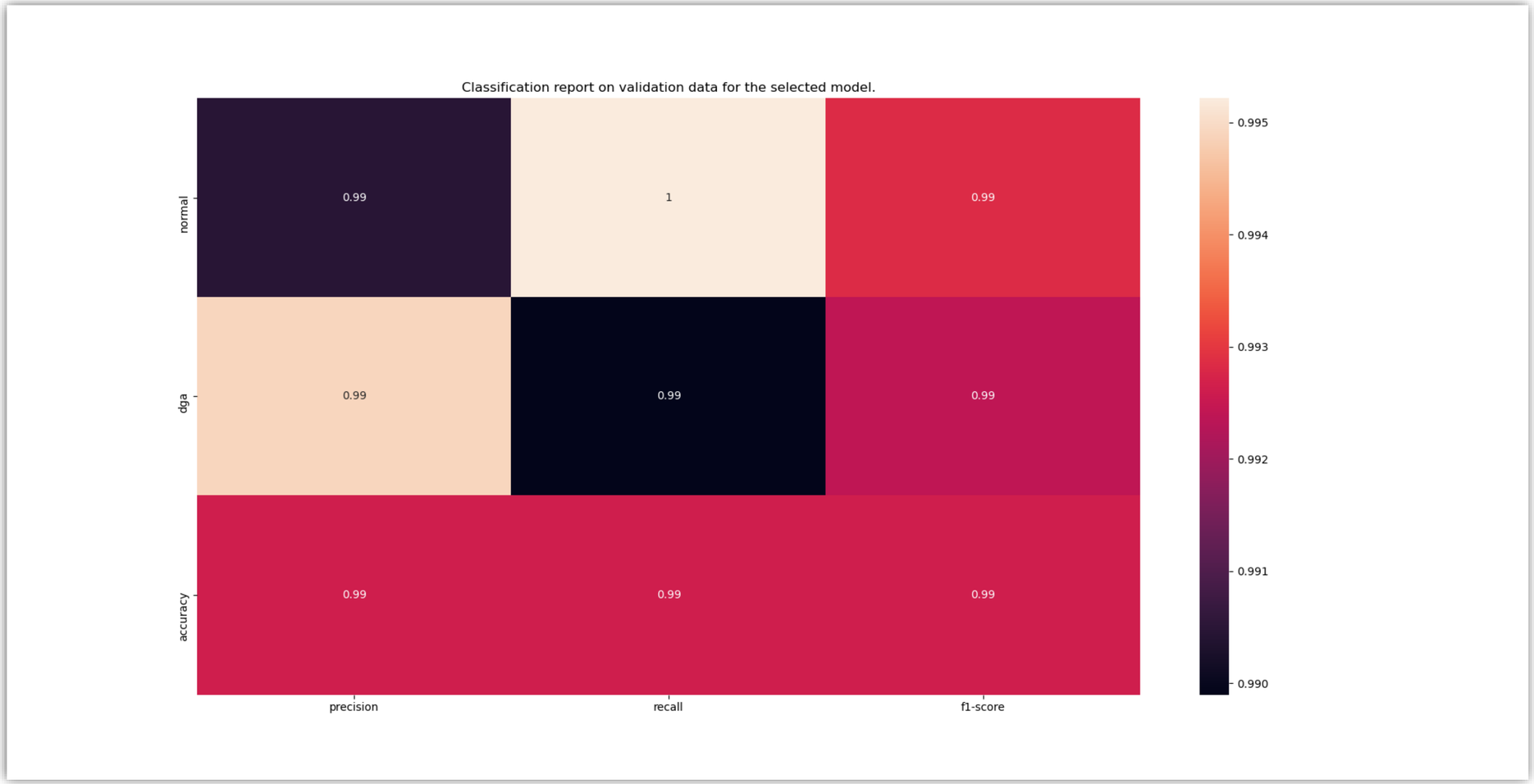
Dataset

- Manual analysis of long term DNS datasets to find representative samples
- 489 IP addresses
  - 240 with proven DGA traffic and different volumes
  - 249 clean DNS traffic different types of clients

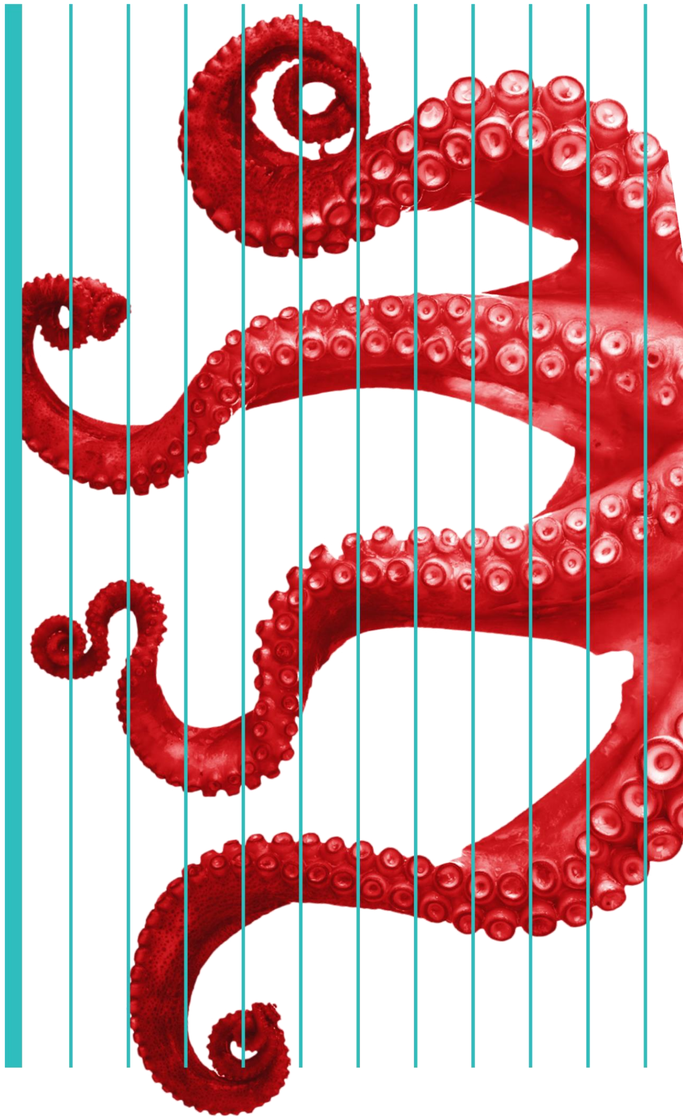
Principal Component Analysis



Principal Component Analysis

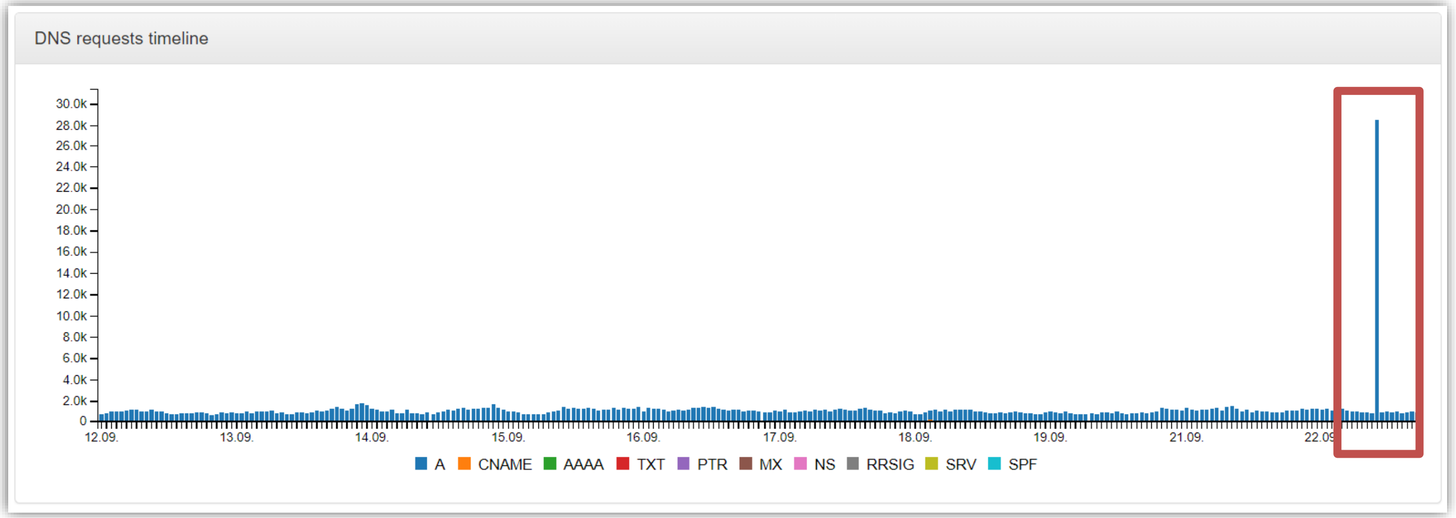


Real world incidents



Monero miner

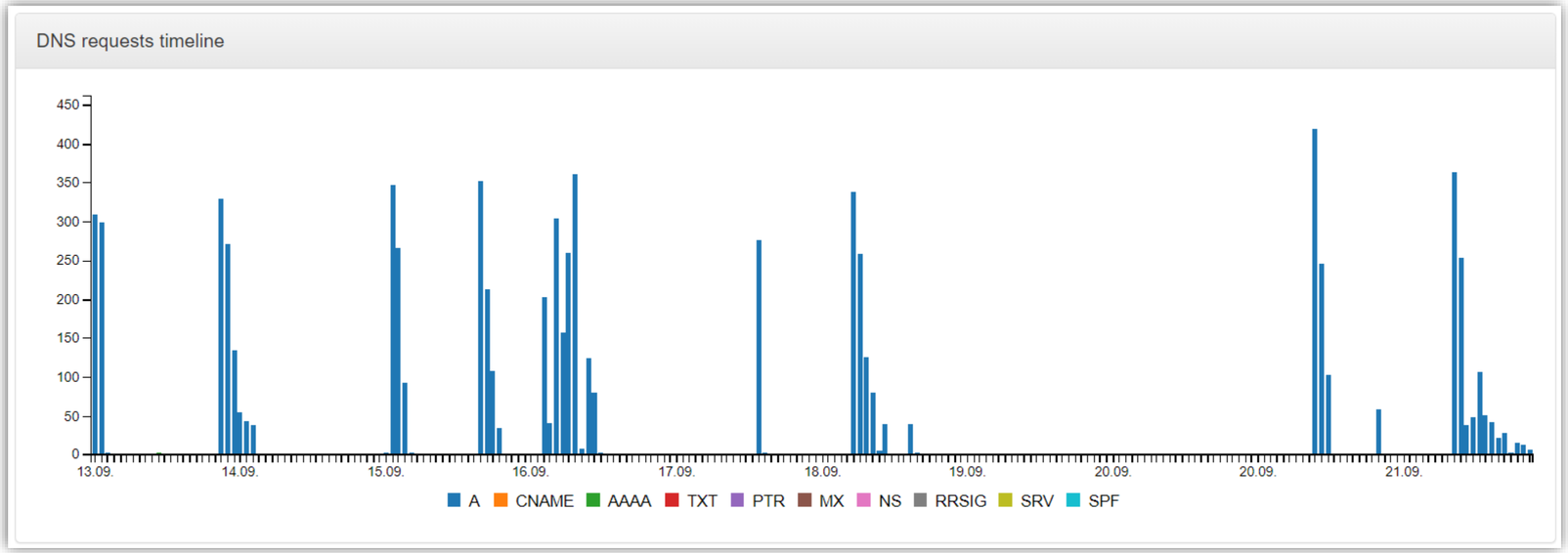
- Thousands of queries per hours
- Combination of TLDs: .hosting, .org, .blackfriday, .tickets, .feedback
- Many active clients for several years



b53caa5c83f5e.hosting
51d585627731c.org
b53caa5c83f5e.blackfriday
cbff63f0cca3e.blackfriday
4288026944a06.feedback
8fce3e9cef5d2.feedback
78501ee9d27ec.tickets
07e1d49d35ca6.blackfriday
3cf058588e842.tickets
90ed4a057711b.tickets
07e1d49d35ca6.hosting
e2dd22b9e6f78.feedback
be2dfd93165b8.tickets
069c4d429610b.org
81974ae793fdc.blackfriday
0e485309af647.tickets
34bc1fcbe9c12.org

Necurs botnet

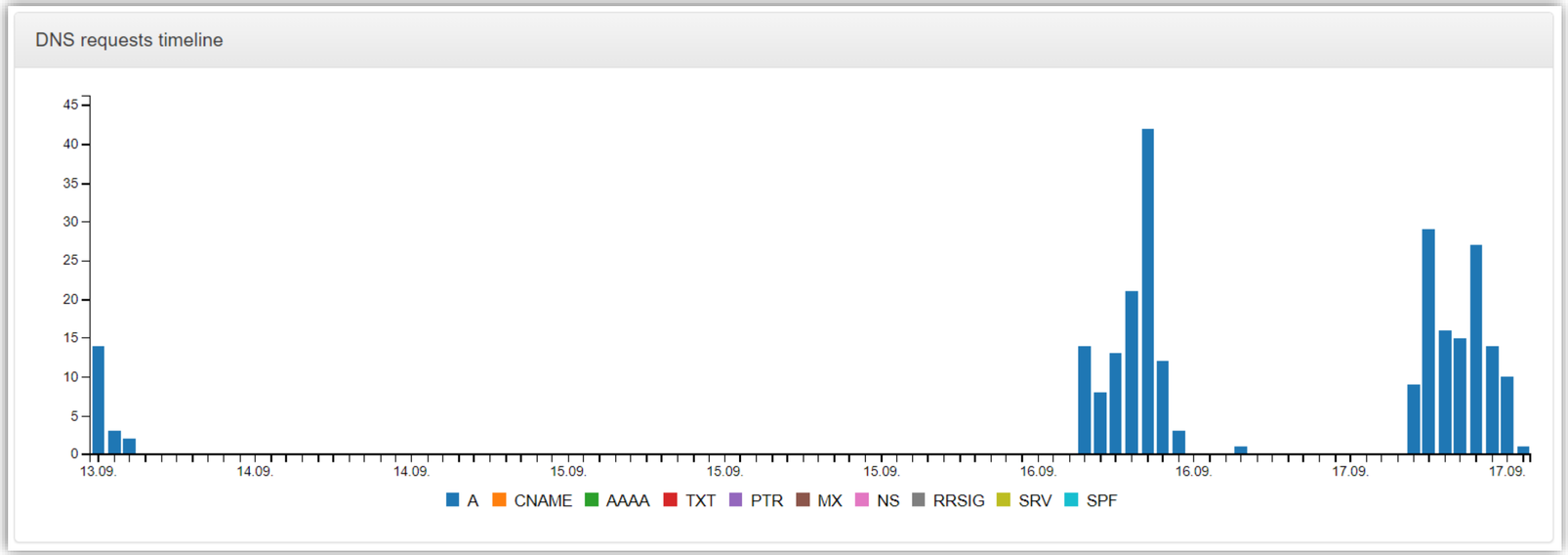
- Hundreds of requests per hour
- Large set of TLDs



hvgkyaxbgyw.mn
wtfltapbyexiyxovnwde.org
elehmysdhewqgjfmfuir.tw
uaytxval.sc
laboynwcigkqu.net
tanzztiikfajarw.com
kpyiljieemnuem.com
qkvxtlafrq.com
gloxiwuethiamvjngmmy.us
frauxatoyqlivyh.sx
esmtedktiktmayqruwh.tw
dldubgnsifapwryrmrxla.mn
tixovijyyijjhmmdpbfi.sx
fwcvahvgv.im
llskqopgsobdgbi.me
sluqoppr.mn
oyuwvkxvgslujrhyiy.sc
agnjshzqvqy.com
iybxeoqjowo.com
rbfzbbsxzxkp.com
dkxflrsfxbvdyek.com
lxnljnyfwhq.com
nsicsvc.sc
kkrrjgmhxuqff.tw
iahdggqddwen.eu

Uknown botnet

- Much more cautious - several queries per hour
- Only TLD is .com

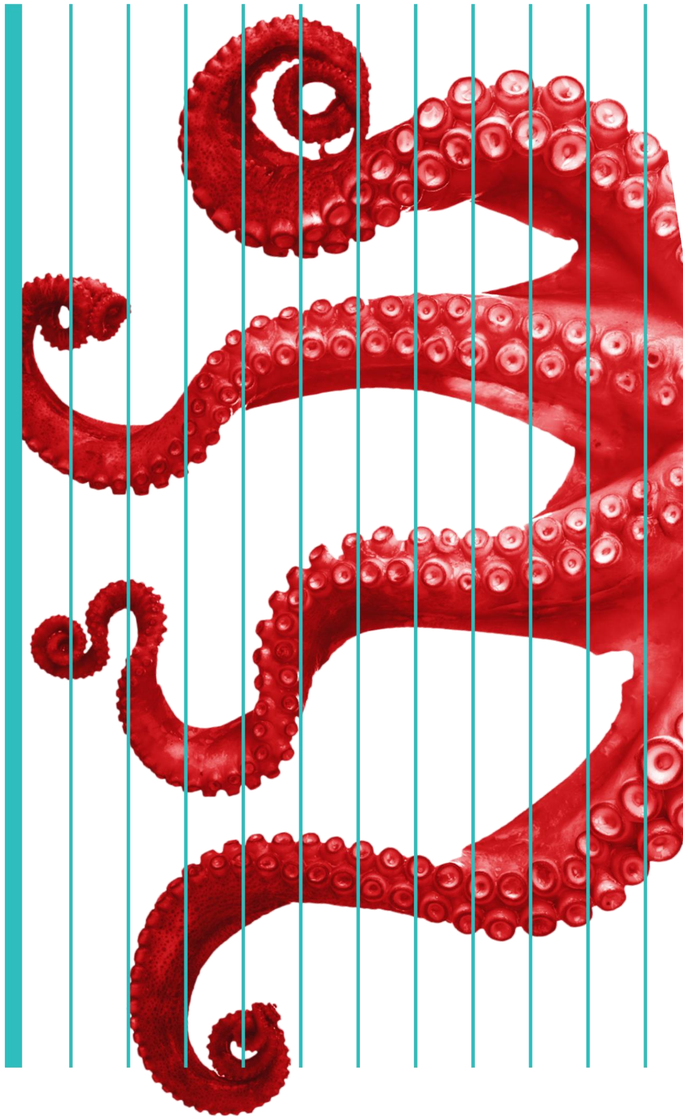


bbskldhxjne.com
tfbcdveeunhlfej.com
dhxdrjhvtxvtyq.com
cfjgccosurkhx.com
vmsbjjrkmmaa.com
cfjgccosurkhx.com
mzgizptbz.com
qpacvjfa.com
bgxvjdrbsgzqxe.com
tqakpsqkywwqlqj.com
idlyyktcpap.com
xdnidssbl.com
uwmqqlnyladot.com
xdnidssbl.com
wedivjci.com
byiffvxobp.com
wqjrgfbdoio.com
byiffvxobp.com
uimgdsrmgdjbdx.com
vidlrlwugwb.com
ixchcnvyu.com
kstgeeebhconmaq.com
mhlexfphol.com

Summary

- Very precise detection based on several hours of data
- Good application neural networks / machine learning
- Simple detection rules could work for obvious or large patterns. but would hold significantly worse True positive / False positive ratio

- Work to do:
  - Real-time prevention (challenging from the performance perspective)



Questions?

robert.sefr@whalebone.io
@robcza