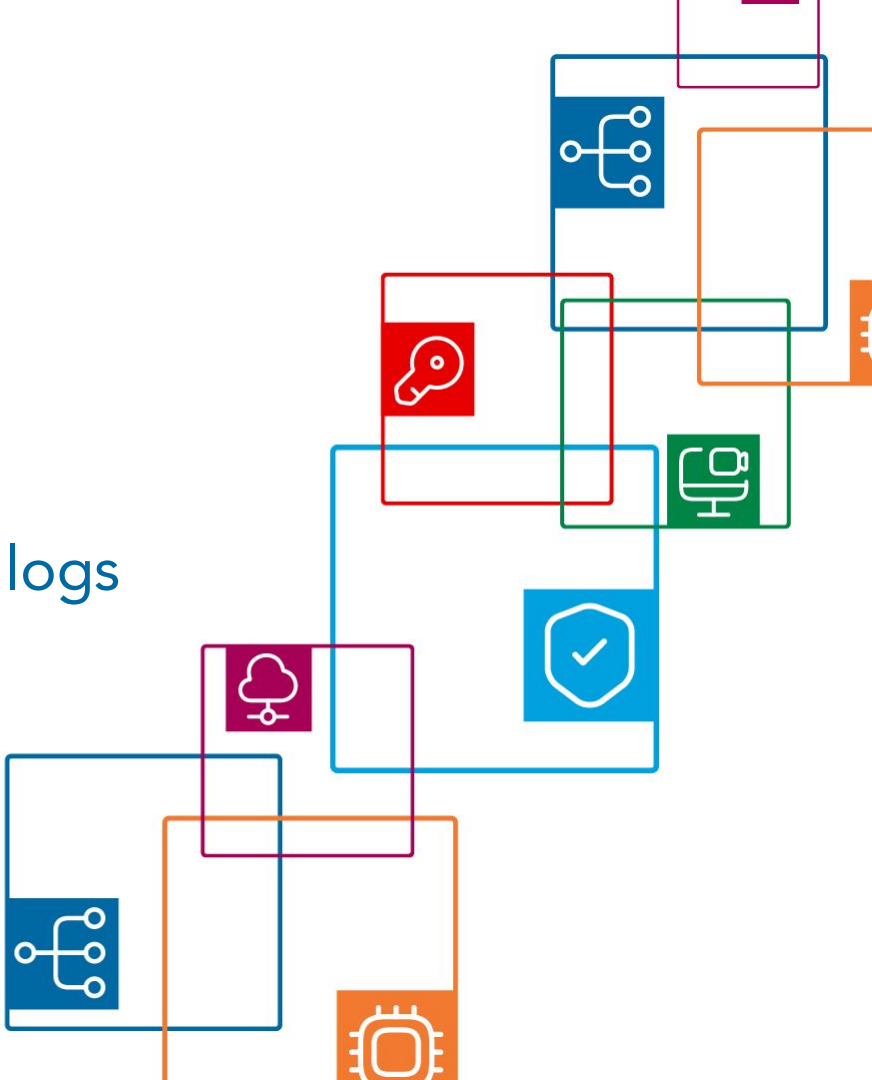
cesnet

DDoS detection from web proxy logs

Jakub Man

April 2026



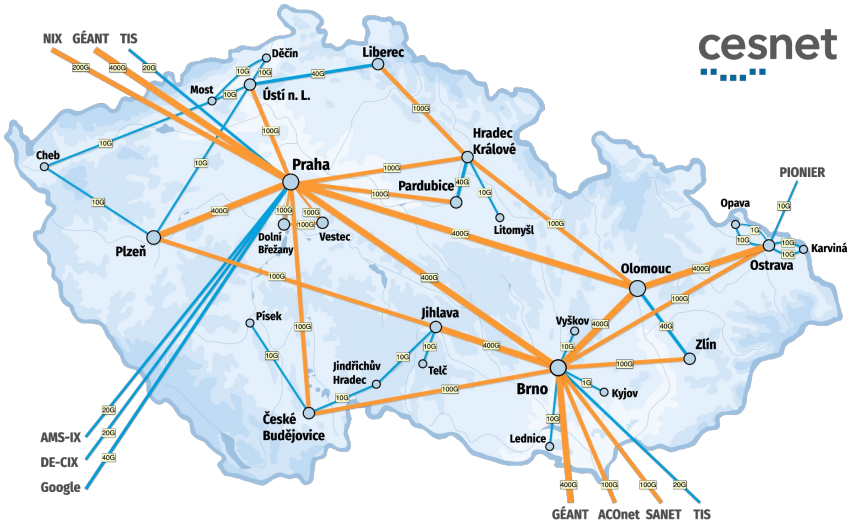
MINISTERSTVO VNITRA
ČESKÉ REPUBLIKY

Created within the project VB02000066, SECTECH, MVČR

cesnet

CESNET

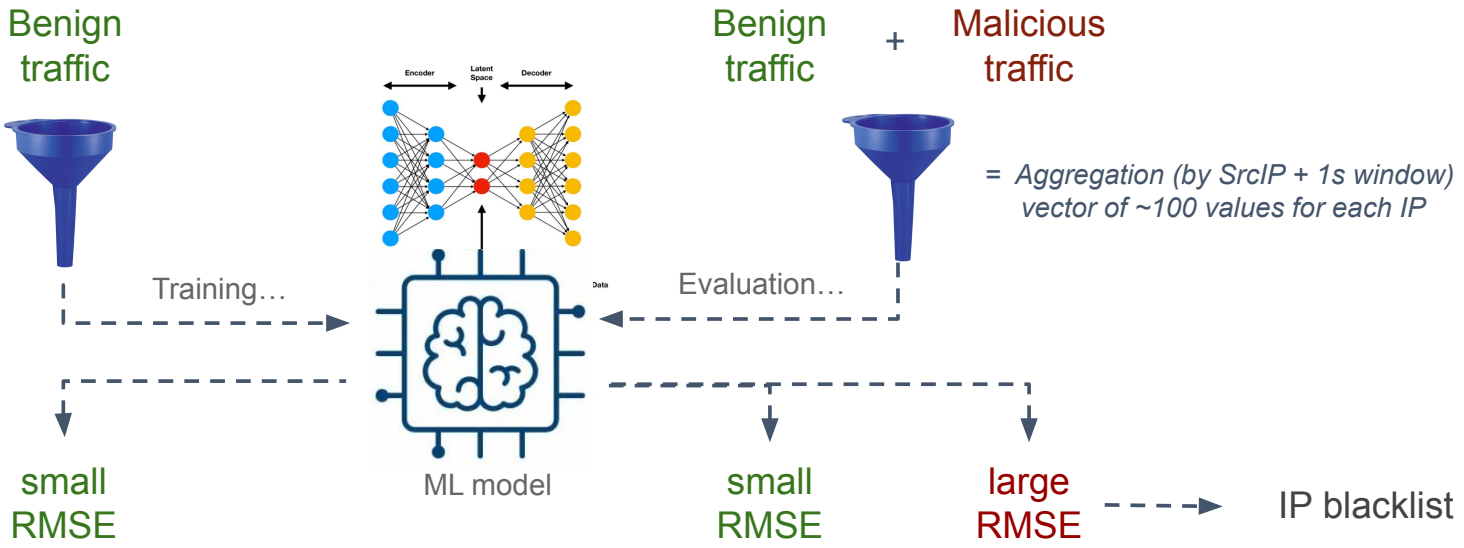
- Association of universities in Czech Republic and the Czech Academy of Science
- Operates and develops e-infrastructure for science, research and education



SEZNAM.CZ

- Leading Czech internet company
- Provides more than 30 products and services such as news, search engine, e-mail, maps, and advertising
- Provided real attack samples for this project and assists with evaluating the results (Thank you!)

Anomalous IP detection - autoencoders

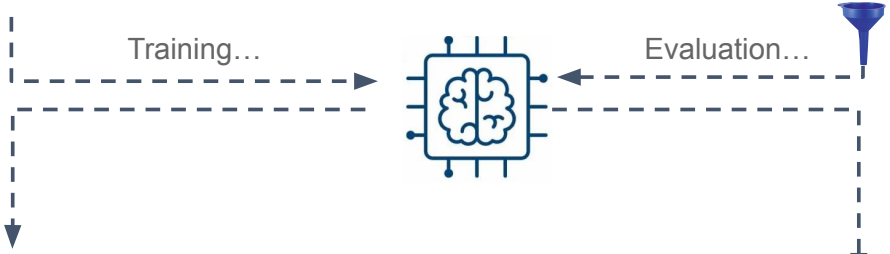


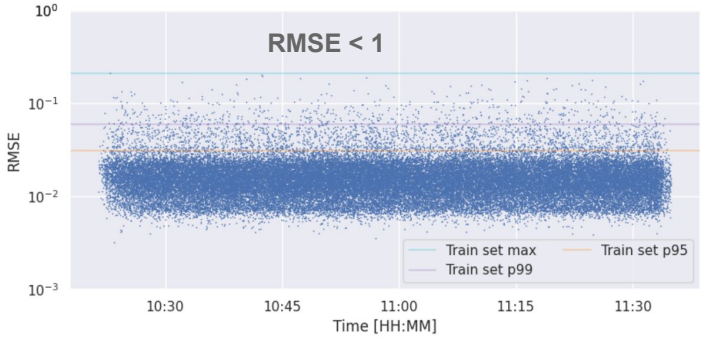
RMSE = Root Mean Square Error

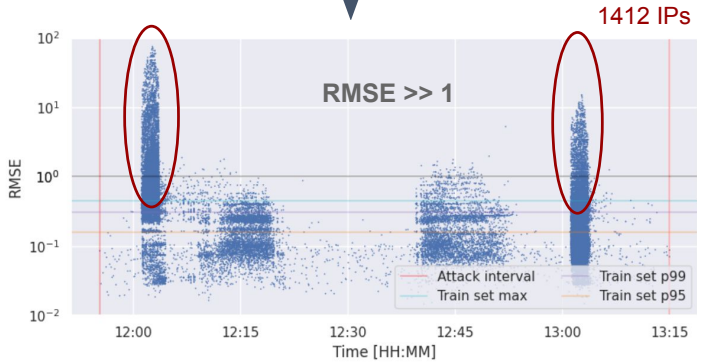
L3/L4 input

L3/L4 benign

L3/L4 benign + malicious





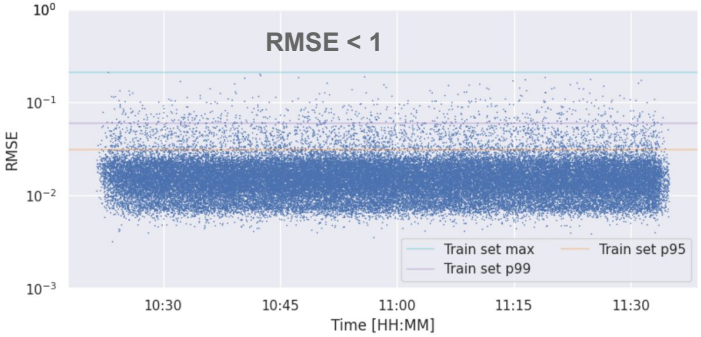


** Each point = IP address evaluation in a specific 1s time window*

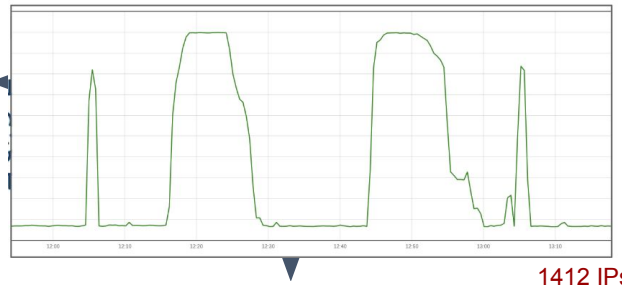
L3/L4 input

L3/L4 benign

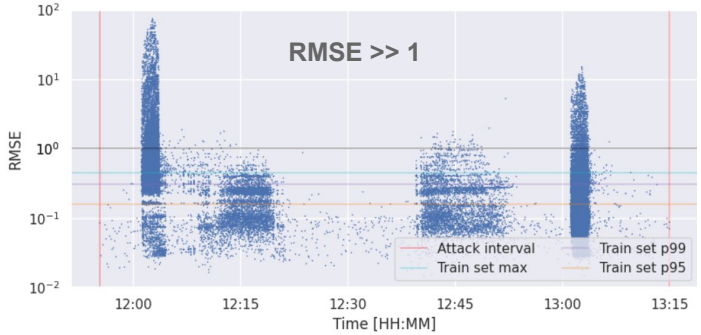
Training…



CPU load during attack



1412 IPs

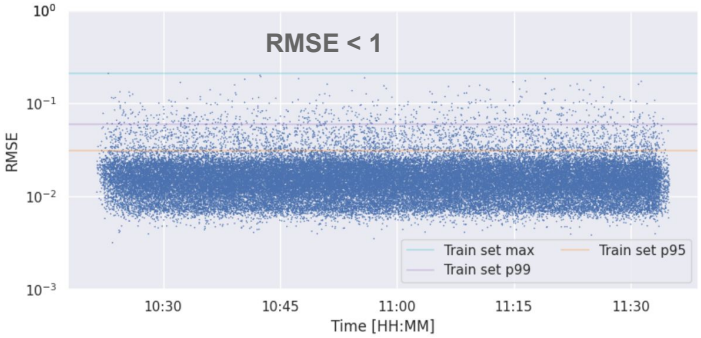


** Each point = IP address evaluation in a specific 1s time window*

L3/L4 input

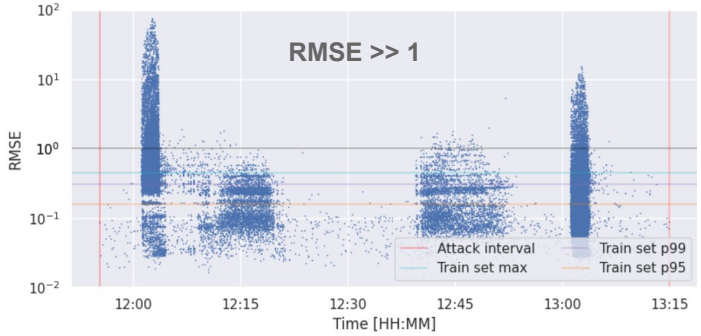
L3/L4 benign

Training…



CPU load during attack

1412 IPs



** Each point = IP address evaluation in a specific 1s time window*

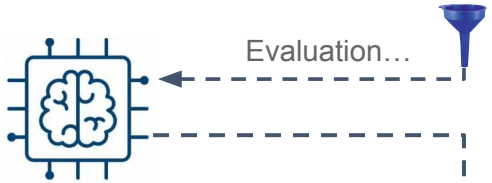
Web proxy

- Central point for all traffic to web services
- Provide traffic logs
- Log fields can be usually configured

L7 input

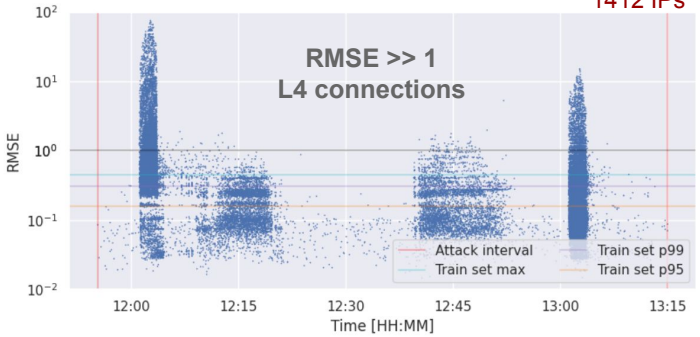
CPU load during attack

L7 logs benign + malicious

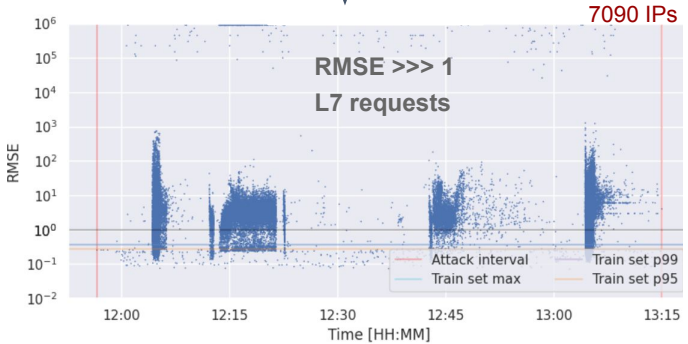


Same features (including aggregation) based on L7 requests instead of packets

1412 IPs

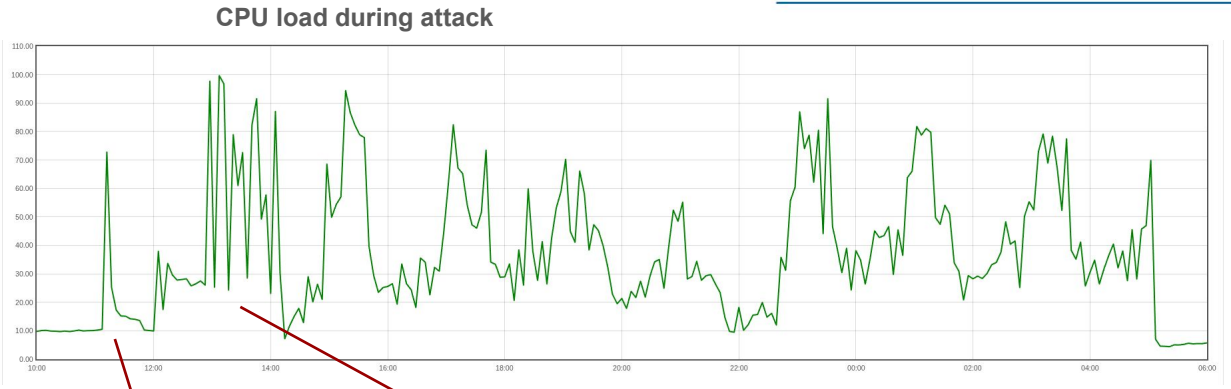


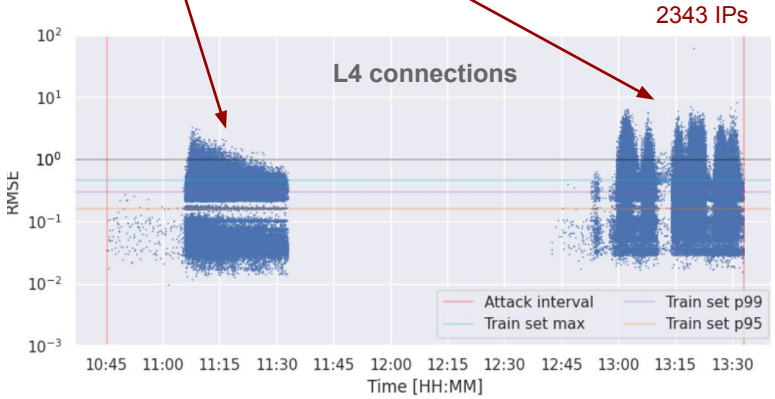
7090 IPs

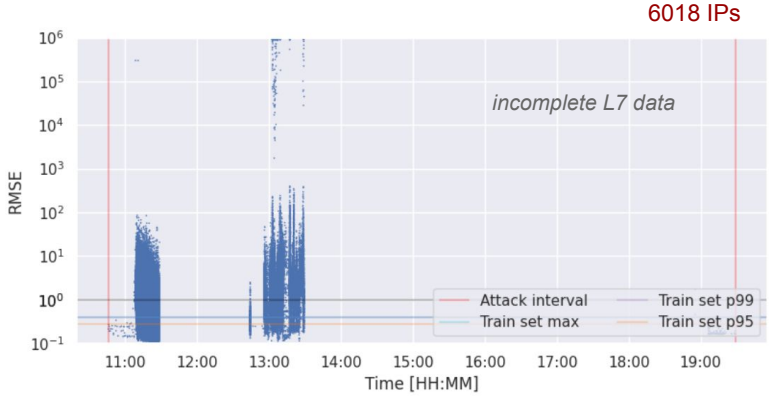


** Each point = IP address evaluation in a specific 1s time window*

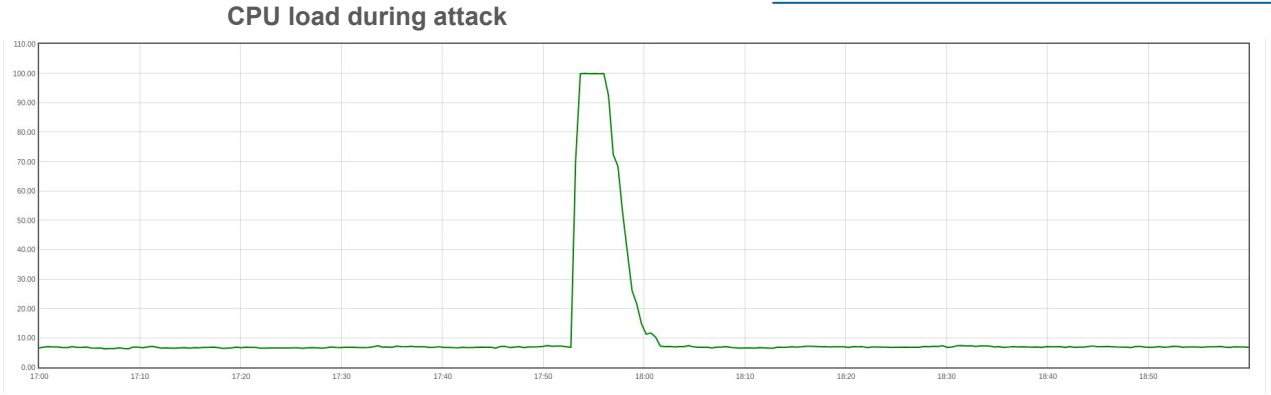
Works on different attack captures



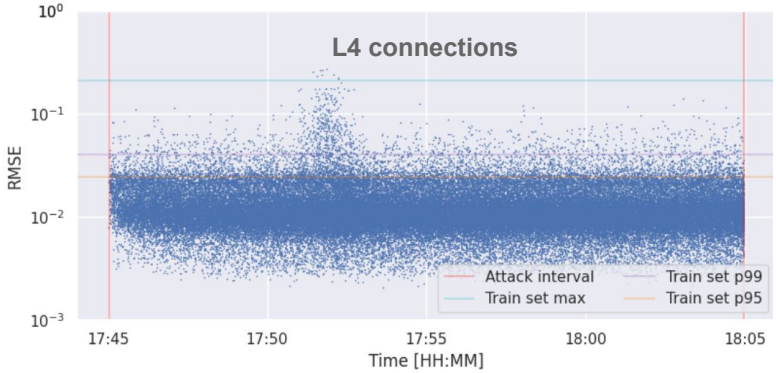




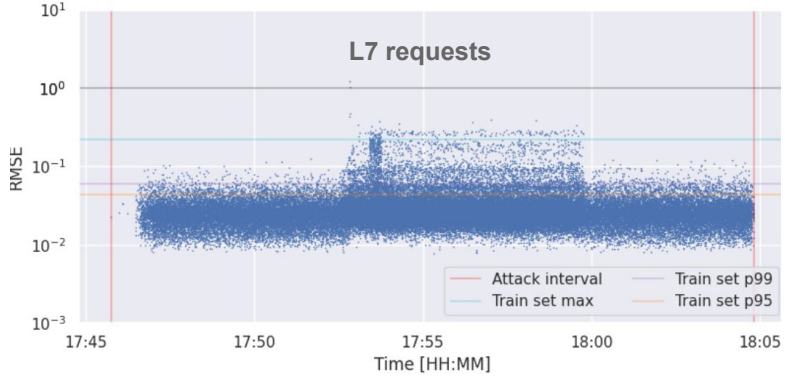
… but not always



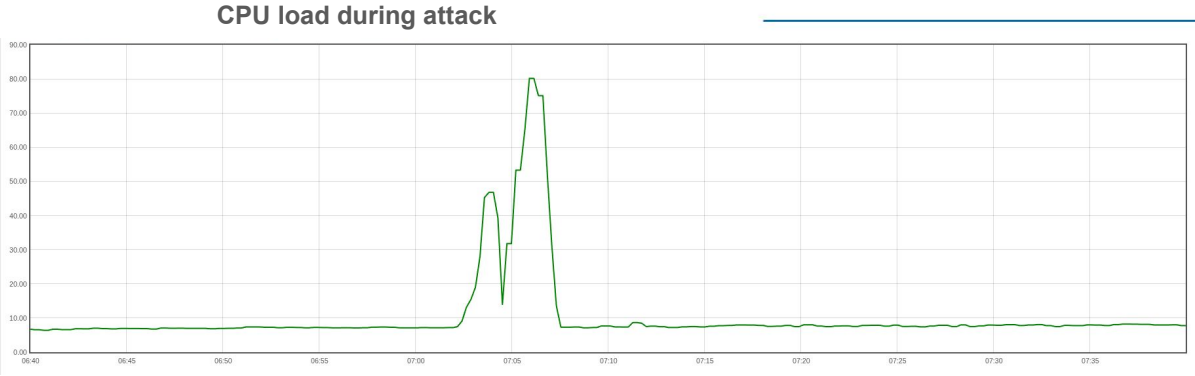
7 IPs



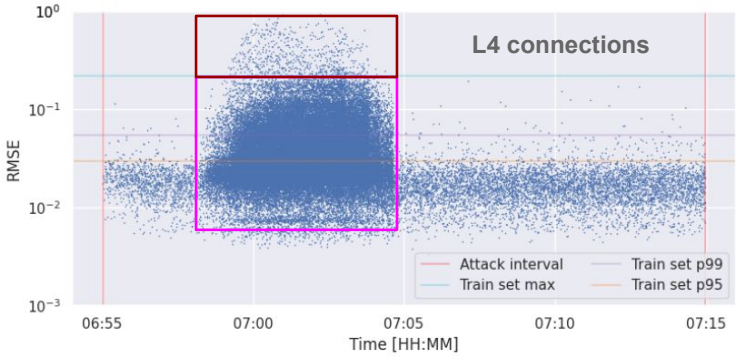
163 IPs



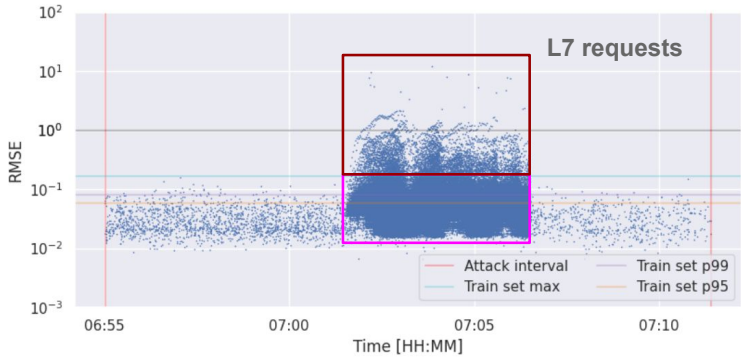
Not anomalous enough



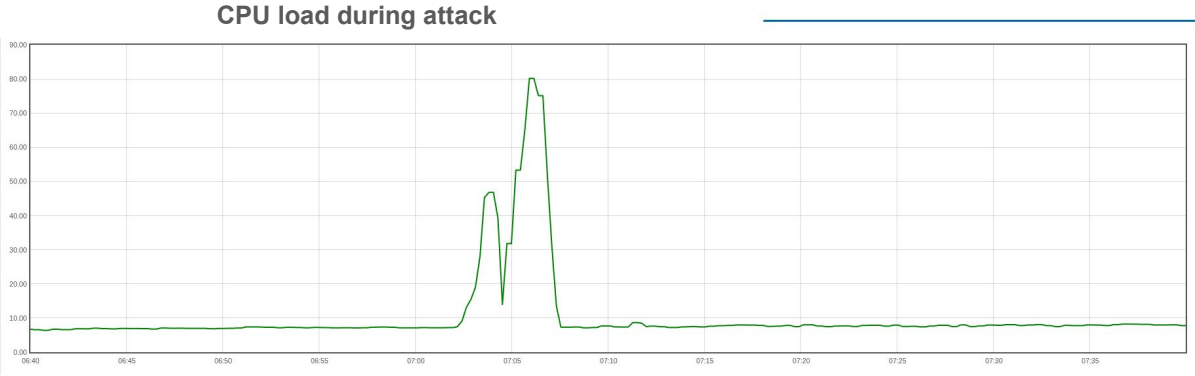
32 IPs

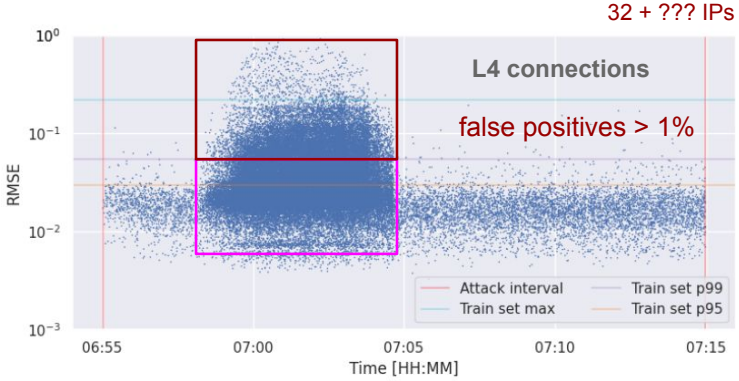


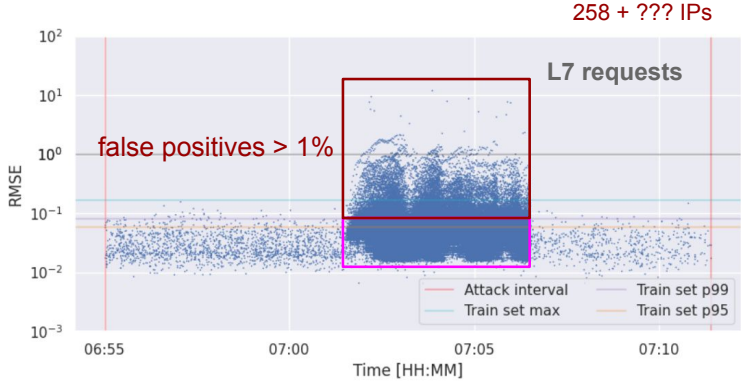
258 IPs



Decrease threshold?

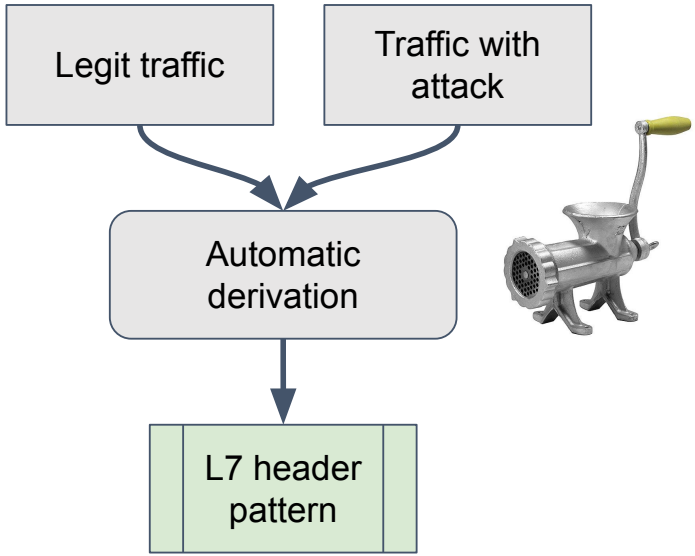






Matching n-grams on L7 logs

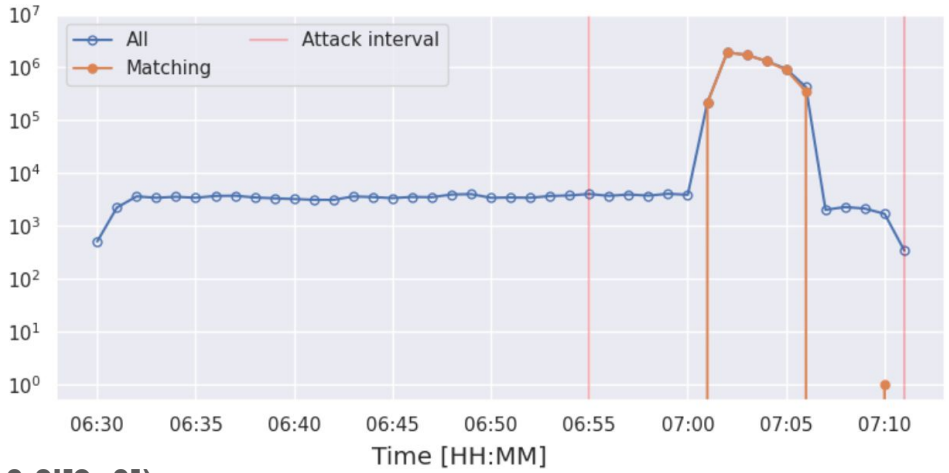
- Input: 2x L7 logs (text / JSON)
- Output: L7 header patterns



cesnet

L7 header patterns

FP = 0.00e+00 (0.00%);
TP = 6.43e+06 (97.31%);



tls_subject: (b'CN=example.com'[0, 0])
tls_cipher_suite: (b'4865'[1, 0])
tls_sni: (b'www.example.com'[2, 0])
tls_version: (b'4'[3, 0])
proto_version: (b'3'[4, 0])
authority: (b'www.example.com'[5, 0])
path: (b'/'[6, 0])
referer: (b'https://www.example.com/'[7, 0])
host: (b'www.example.com'[8, 0])
accept_lang: (b'es-AR,es;q=0.8,en-US;q=0.5,en;q=0.3'[9, 0])
accept_enc: (b'gzip, deflate, br'[11, 0])
accept: (b'text/html,application/xhtml+xml,application/xml;q=0.9,image/avif,image/webp,*/*;q=0.8'[13, 0])
method: (b'1'[14, 0])
scheme: (b'https'[15, 0])
user_agent: (b'Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)'[16, 0])
rc: (b'0'[17, 1])

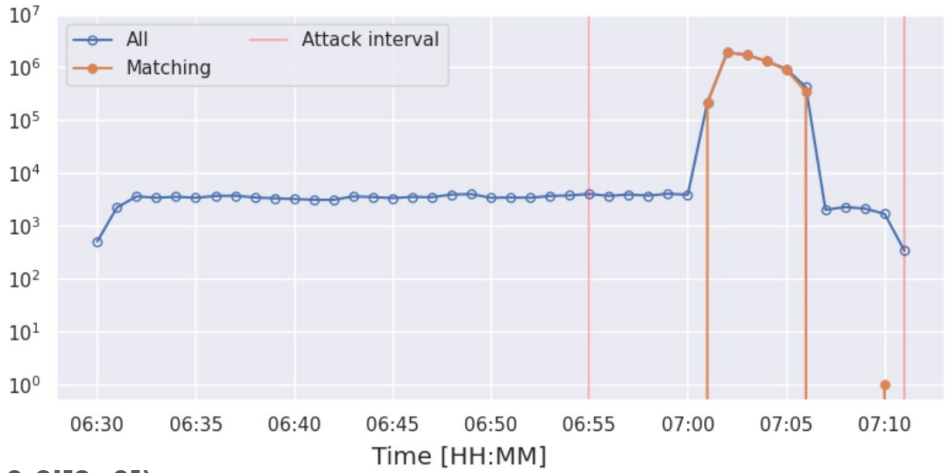
{"0": "tls_subject", "1": "tls_cipher_suite", "2": "tls_sni", "3": "tls_version", "4": "proto_version", "5": "authority", "6": "path", "7": "referer", "8": "host", "9": "accept_lang", "10": "accept_dt", "11": "accept_enc", "12": "connection", "13": "accept", "14": "method", "15": "scheme", "16": "user_agent", "17": "rc"}

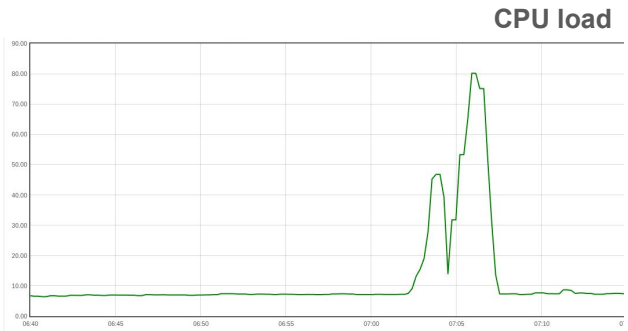
cesnet

L7 header patterns

FP = 0.00e+00 (0.00%);
TP = 6.43e+06 (97.31%);

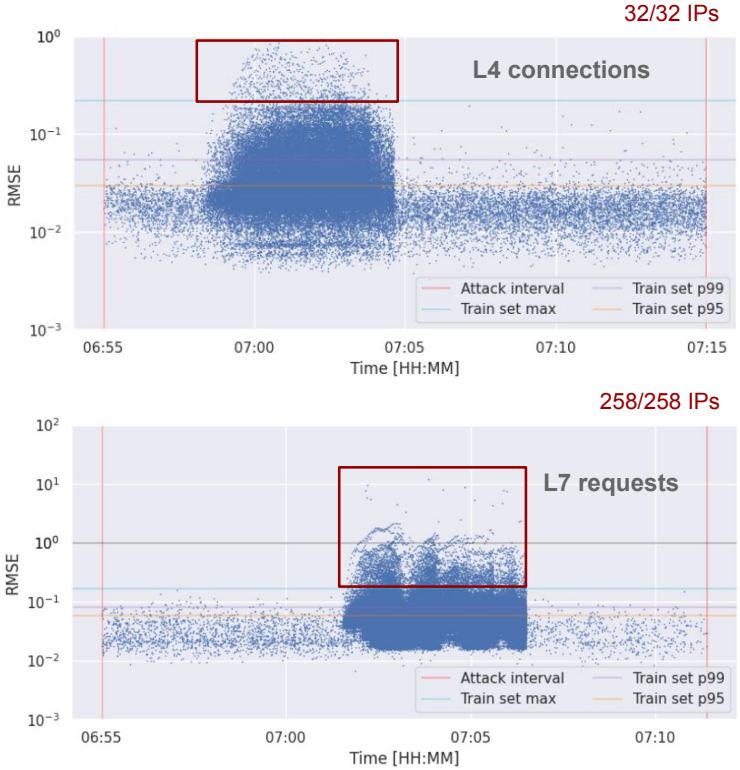
tls_subject: (b'CN=example.com'[0, 0])
tls_cipher_suite: (b'4865'[1, 0])
tls_sni: (b'www.example.com'[2, 0])
tls_version: (b'4'[3, 0])
proto_version: (b'3'[4, 0])
authority: (b'www.example.com'[5, 0])
path: (b'/'[6, 0])
referer: (b'https://www.example.com/'[7, 0])
host: (b'www.example.com'[8, 0])
accept_lang: (b'es-AR,es;q=0.8,en-US;q=0.5,en;q=0.3'[9, 0])
accept_enc: (b'gzip, deflate, br'[11, 0])
accept: (b'text/html,application/xhtml+xml,application/xml;q=0.9,imag
method: (b'1'[14, 0])
scheme: (b'https'[15, 0])
user_agent: (b'Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.go
rc: (b'0'[17, 1])

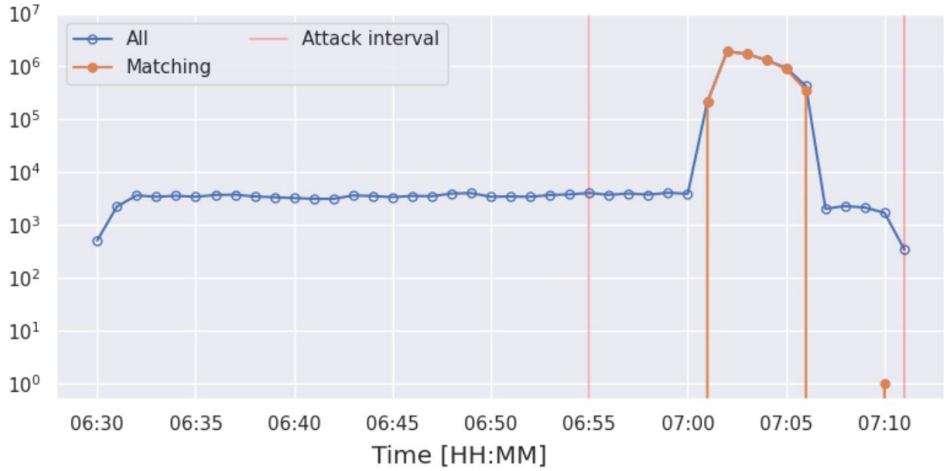


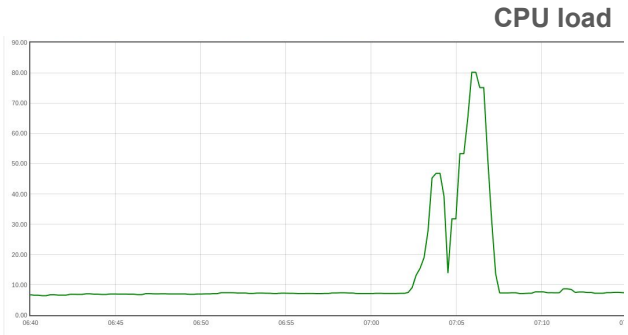


{"0": "tls_subject", "1": "tls_cipher_suite", "2": "tls_sni", "3": "tls_version", "4": "proto_version", "5
"9": "accept_lang", "10": "accept_dt", "11": "accept_enc", "12": "connection", "13": "accept", "14": "method", "15": "scheme", "16": "user_agent", "17": "rc"}

L7 header patterns



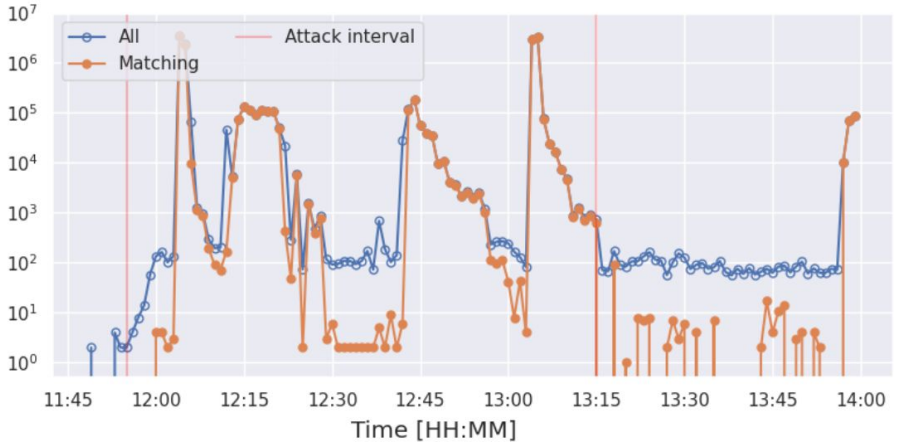




L7 patterns - first dataset

FP = 0.00e+00 (0.00%);
TP = 1.35e+07 (97.41%);

tls_subject: (b'CN=a'[0, 0])
tls_cipher_suite: (b'4865'[1, 0])
tls_sni: (b'a'[2, 0] && b'.e'[2, 6])
tls_version: (b'4'[3, 0])
proto_version: (b'3'[4, 0])
authority: (b'a'[5, 0] && b'.e'[5, 6])
path: (b'/'[6, 0])
referer: (b'a'[8, 0] && b'.e'[8, 5])
accept_lang: (b'-'[9, 2] && b','[9, 5] && b';q=0.'[9, 8])
accept_enc: (b'gzip, deflate, br'[11, 0])
scheme: (b'https'[15, 0])

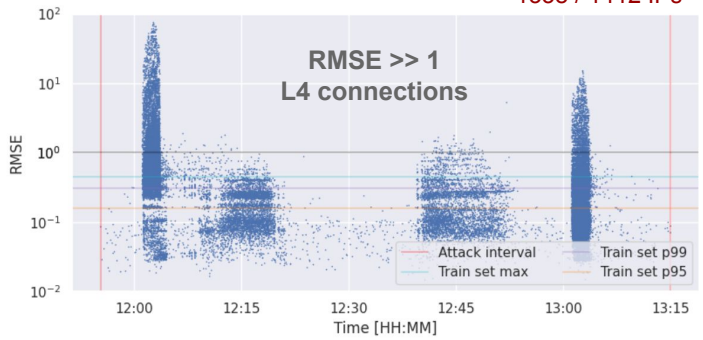


Examples:

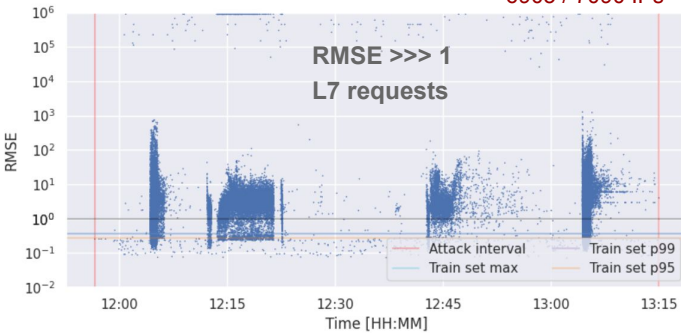
tls_subject: CN=alpha.example.com | CN=admin.elpmaxe.com
tls_sni: alpha.example.com | admin.elpmaxe.com
accept_lang: ru-RU,ru;q=0.9,en-US;q=0.8,en;q=0.7 | en-US,en;q=0.9,ru;q=0.8,de;q=0.7

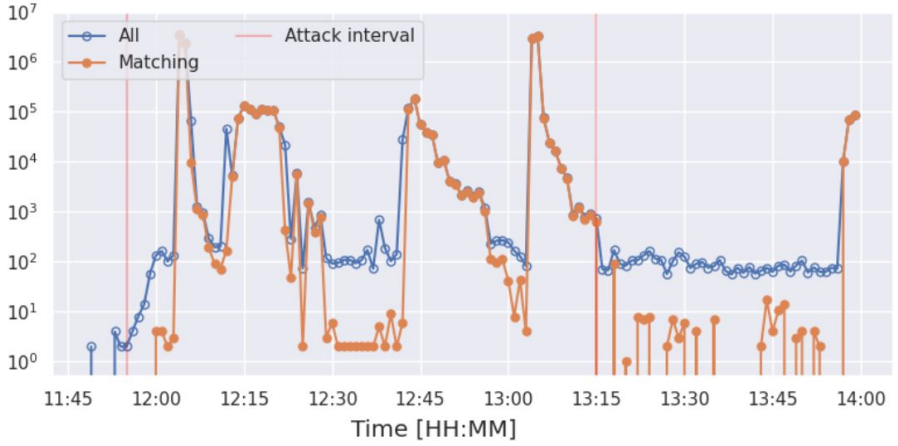
L7 patterns - first dataset

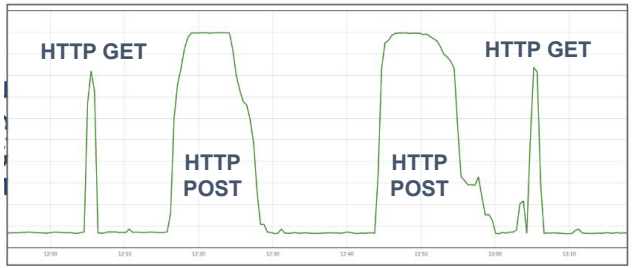
1338 / 1412 IPs



6905 / 7090 IPs







cesnet

Algorithm comparison

Autoencoders (L4)

- Works on encrypted traffic
- Needs to parse packets - slower than L7 version

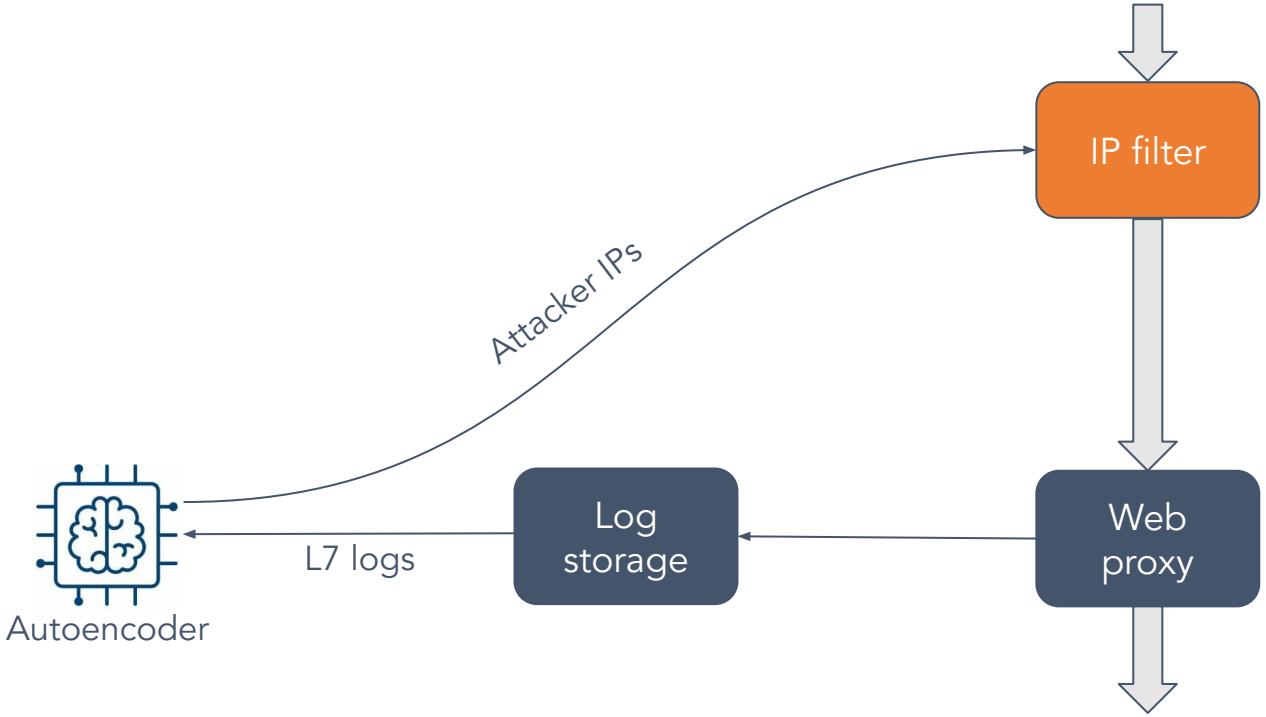
N-gram patterns

- Too slow to process all traffic
- Outputs precise filters
- Human-explainable output

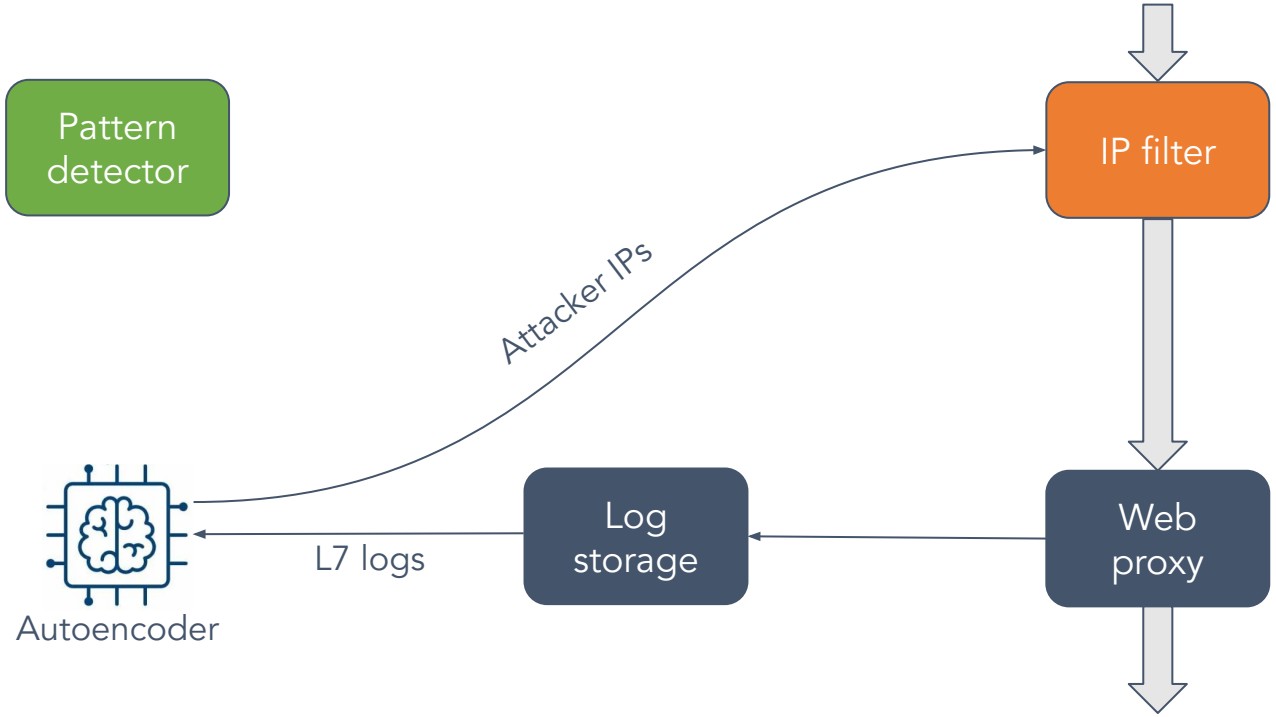
Autoencoders (L7)

- Fast enough to observe all live traffic (using L7 logs)
- Might miss some addresses with low RSME

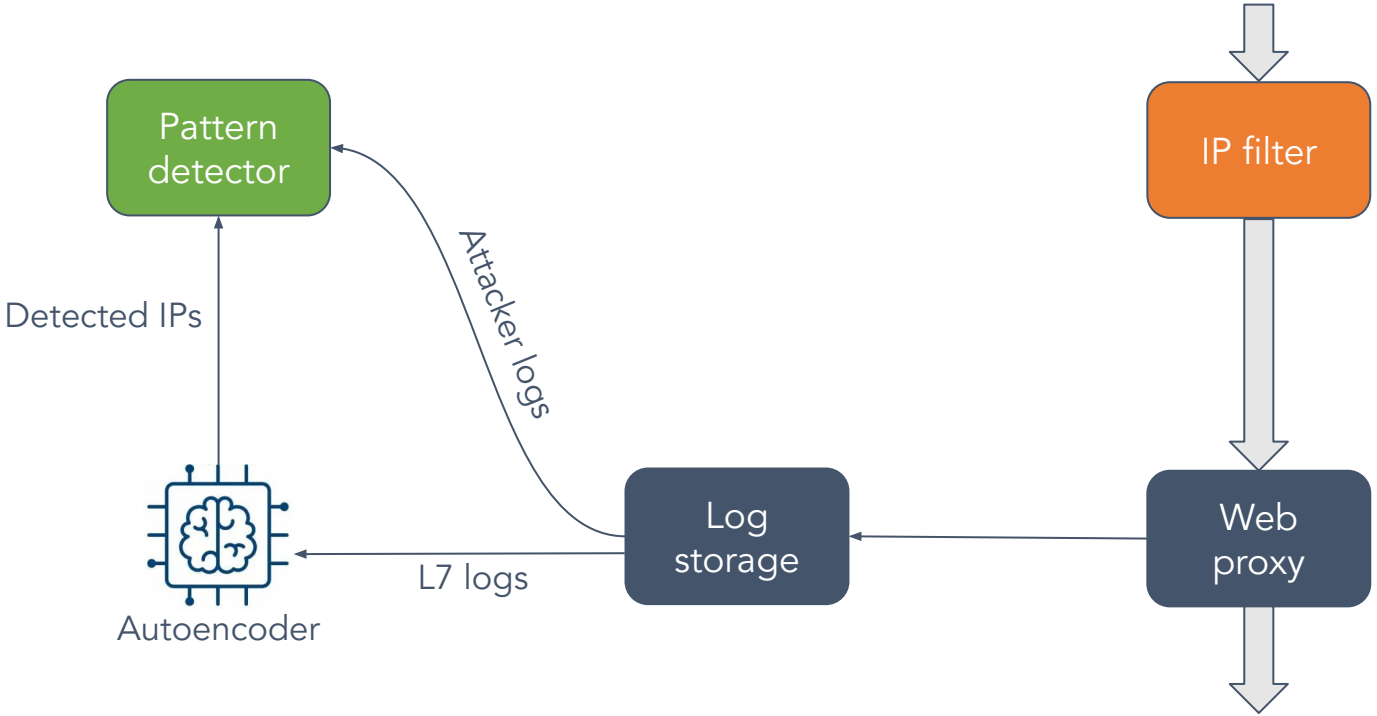
Autoencoder deployment



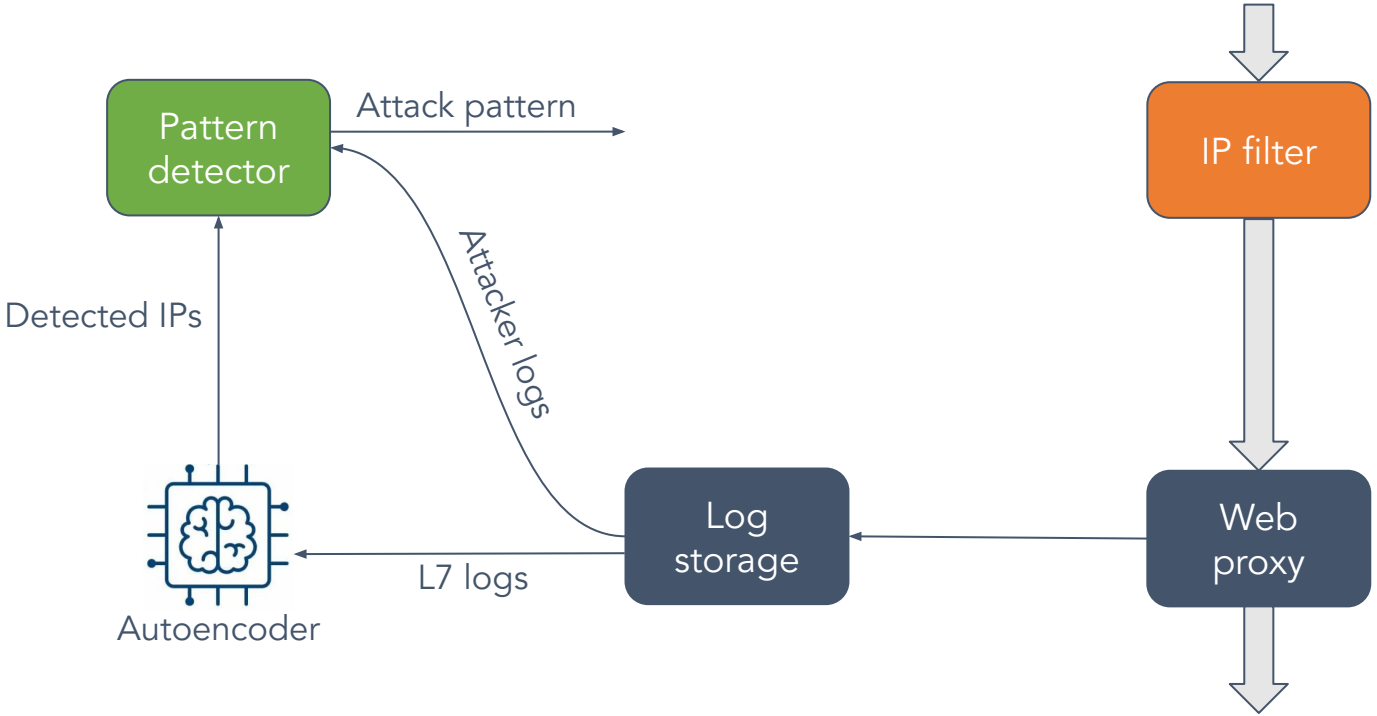
Combined deployment



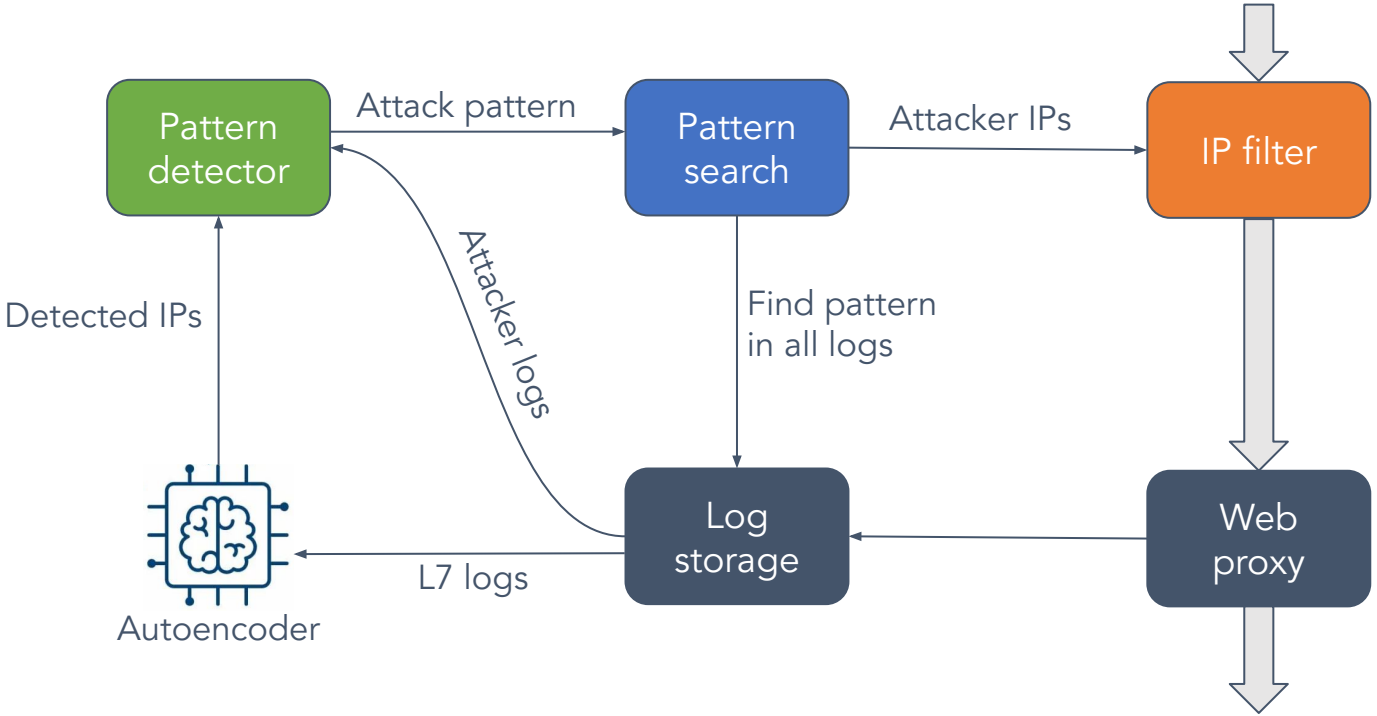
Combined deployment



Combined deployment



Combined deployment



cesnet

Further work

- Improve pattern detection performance
- Tooling for integration into networks
- Experiment with using timing statistics for attack detection

cesnet

Thank you

ddp@cesnet.cz

Jakub.Man@cesnet.cz

liberouter.org/ddos-protector