cesnet

Overview of selected AI projects at CESNET

Karel Hynek, Martin Žádník
CESNET

2025

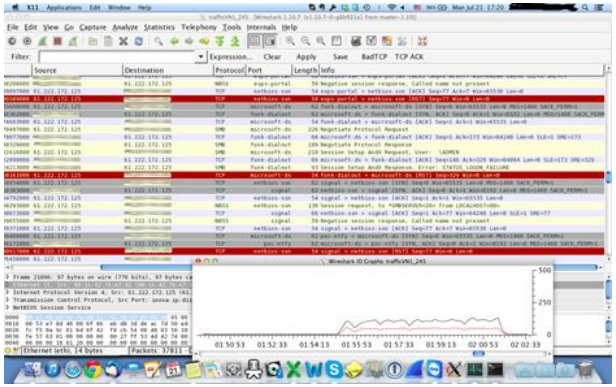
- Research of prospective SOC/NOC cybersecurity tasks within the incident handling lifecycle

cesnet

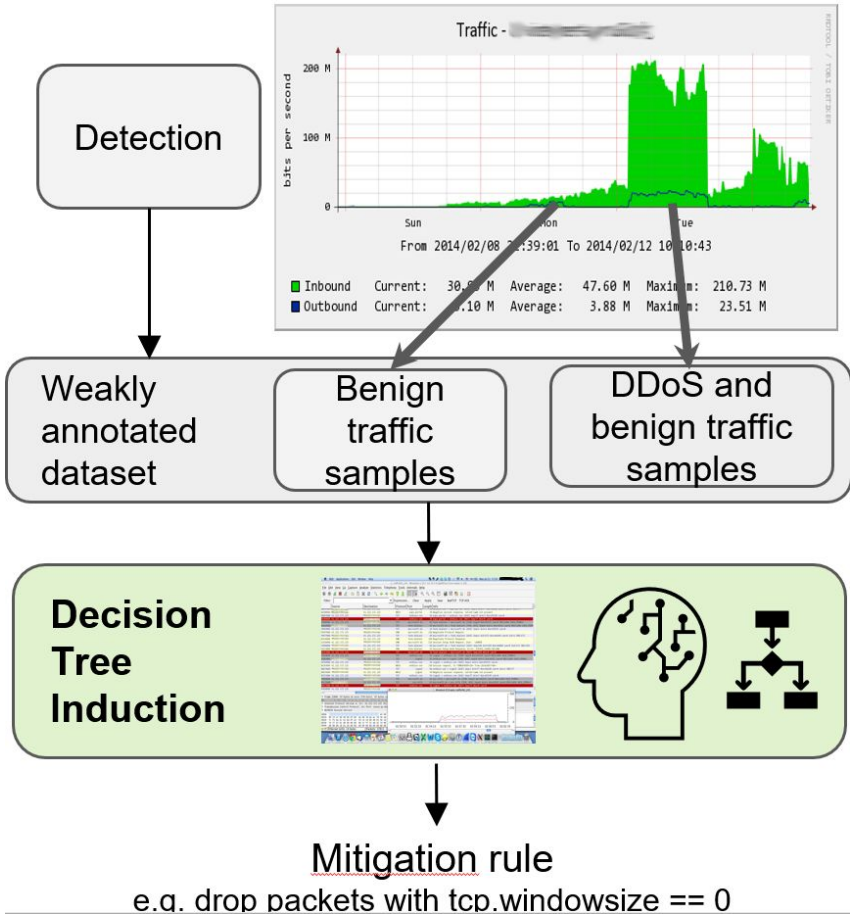
ML mitigates DDoS attacks

Inference of DDoS mitigation rules

- CESNET develops its DDoS protection
- DDoS attacks are short, multivector
- Lack of human resources to respond
- Fast automated response to support human



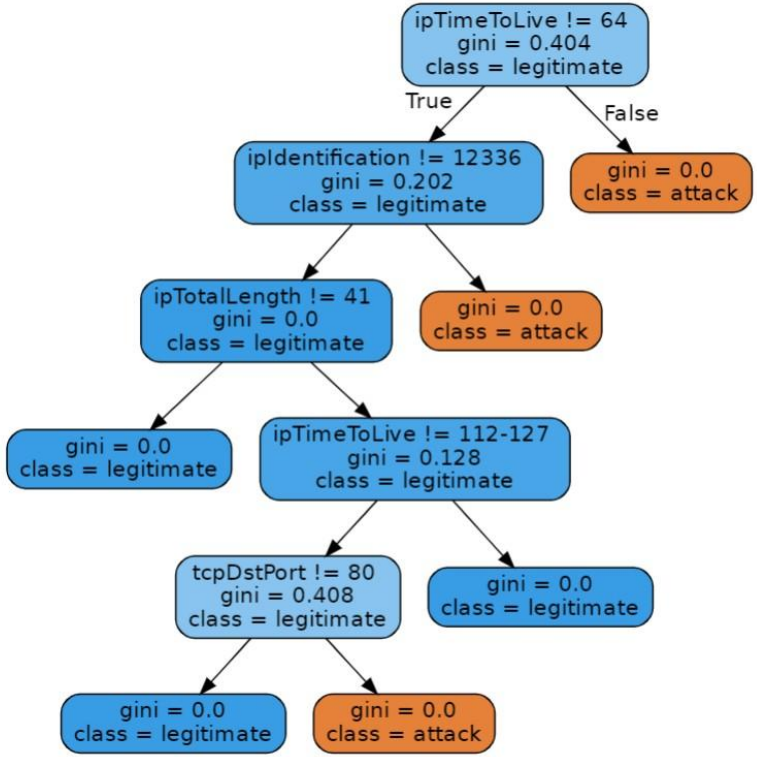
Inference of DDoS mitigation rules



- Assumptions for ML to work with poor annotation
  - DDoS packets represents the majority of the traffic mix during attack
  - DDoS packets are similar to each other considering some packet header fields

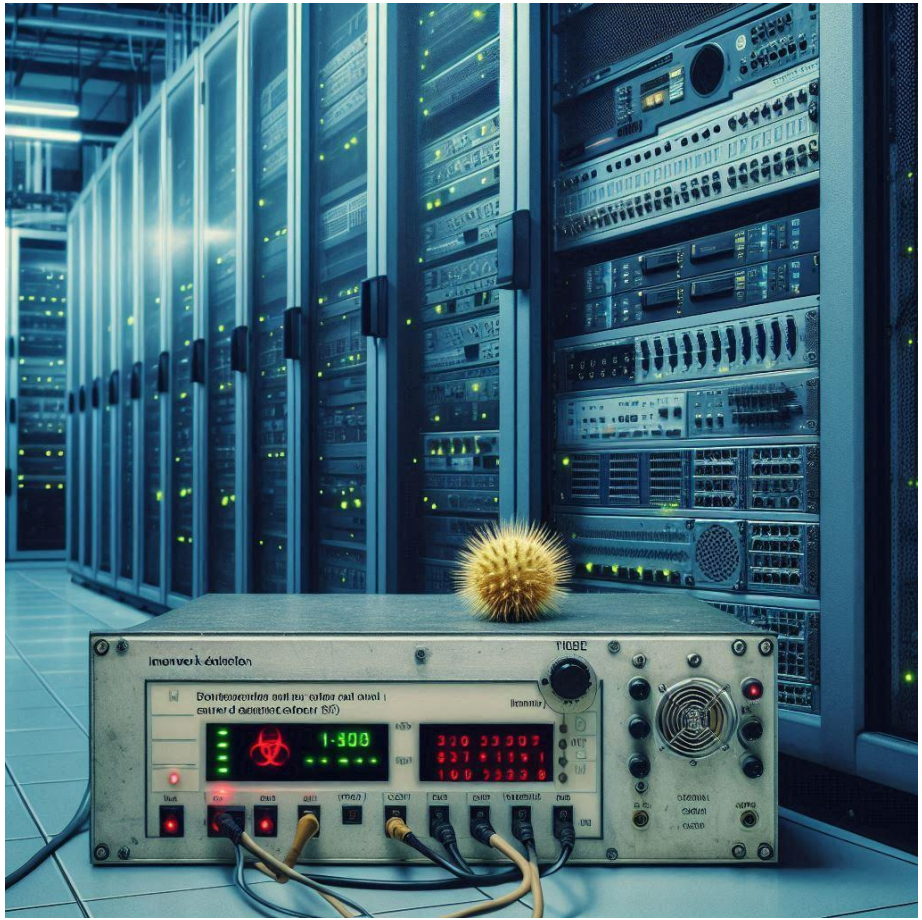
Inference of DDoS mitigation rules

- Example
  - Tree and respective rules
  - multivector DDoS

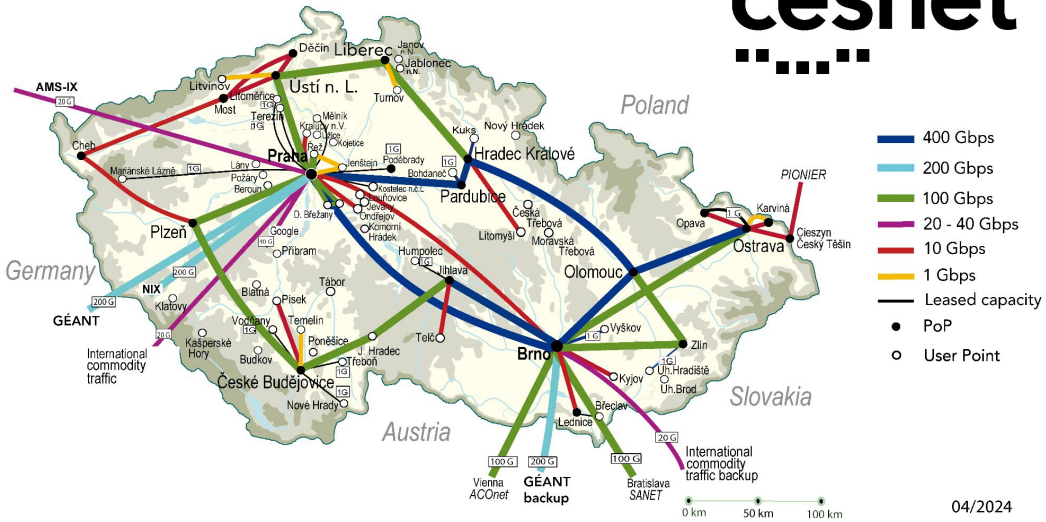



```
(ipTTL == 64) or (ipID == 12336) or
(ipLEN == 41 and ipTTL !=112-127 and
tcpDPORT == 80)
```

Traffic Analysis and Threat detection



Network monitoring

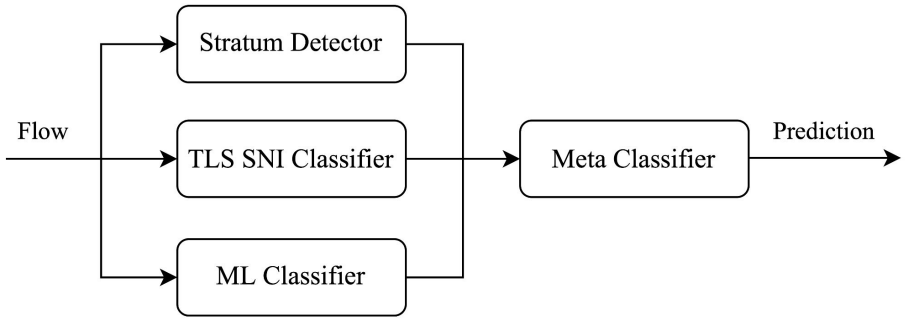


ipfixprobe

- Highly enriched bidirectional flow data
  - IP addresses, number of packets, number of bytes, TCP flags…
  - 100B of payload
  - Sequence of packet lengths and times (30 packets)
  - TLS SNI, QUIC SNI, JA3…

Miner Detector

- CESNET operates MetaCentrum – Large research computational grid
  - It should be protected from misuse—GPUs should not be used for mining!



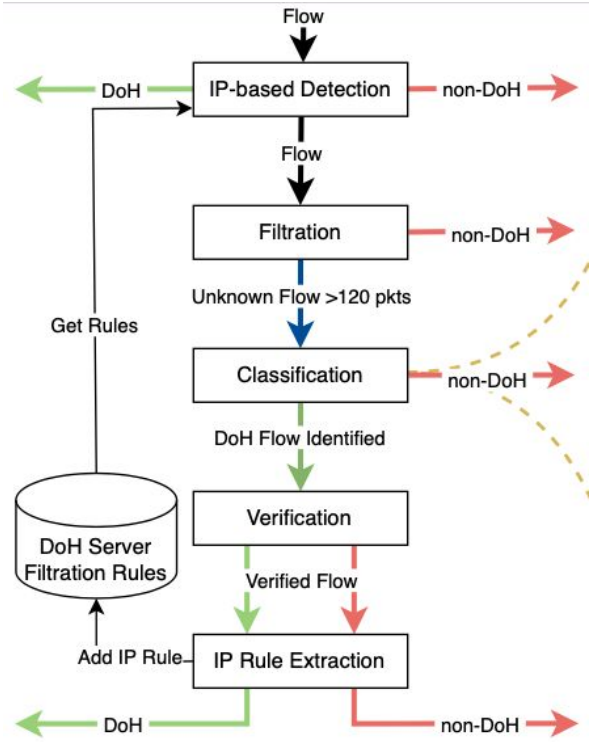
- There is minimal number of false positives due to heterogeneity of the classifier
- Provides explainable alerts
  - SNI: pool.eshop
  - ML: Crypto Mining Pool

SSH Brute-Force Detector

- Packet sequence allow us to recognize successful/unsuccessful login despite the encryption
  - Packet indicating successful login is much larger than unsuccessful
- AI is used to determine encryption algorithm from sizes of packets
  - making the recognition much more accurate
- Uses data from whole CESNET network
  - Able to recognize even low&slow bruteforce attacks

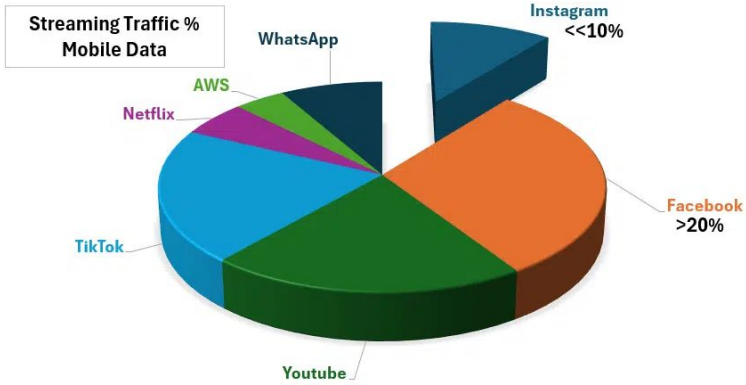
AI used as prefilter for DoH server identification

- Indication of DoH is necessary for IDP/IPS
  - It is strange when client connects to IP that was not queried via DNS
- AI classifier is not performing final detection, just selects highly probable candidates
- The final classification is based on external API
  - AI prevents overloading the external API
- In case of DoH, almost none false negatives are found
  - The ML model is tuned to create more false positives, that are then filtered out

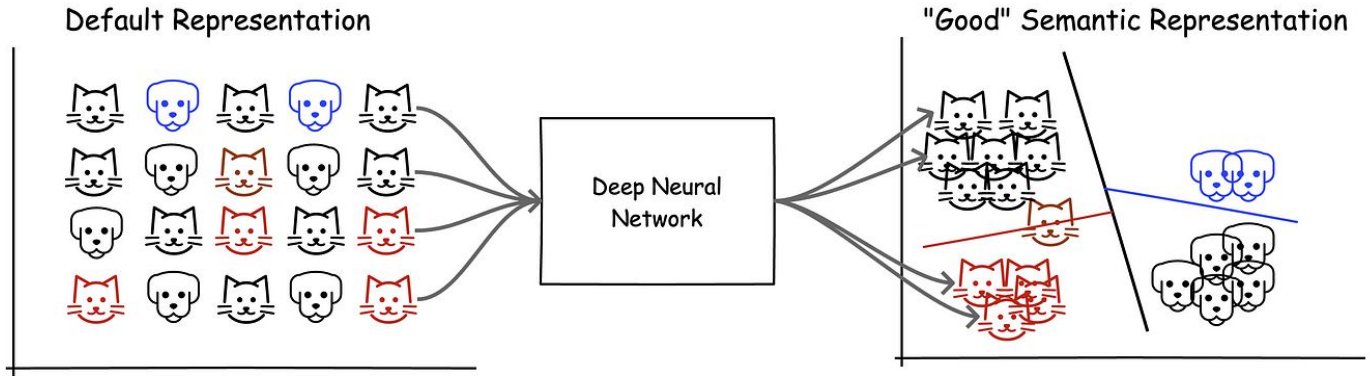


Domain categorization - LLM as a automated Google API

- Preparation of dashboard with types of traffic
  - Video, social, news…
- It is usually based on lists of domain names with corresponding category
- New domain name→ google search → first result → category
- LLM for domain categorization into services
  - it effectively servers as automated google API



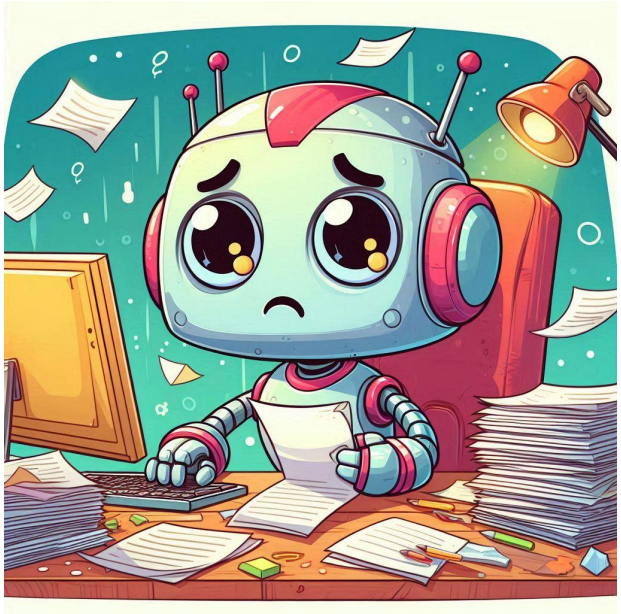
Research in Metric Learning



- Pulling similar input sample to the same vector representation
- NN learns important features and discards the noise
- Similarity-based classification
  - If observed communication is similar to communication with blacklisted IP
  - New IP can be added to blacklist
  - If too many communications are similar to communication with blacklisted IP—it is probably FP

cesnet

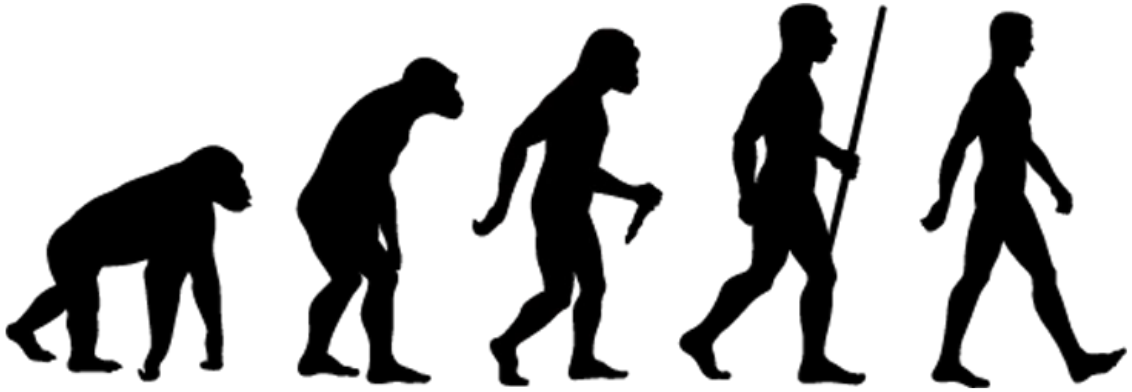
AI is always used as support!
Why not use AI directly?



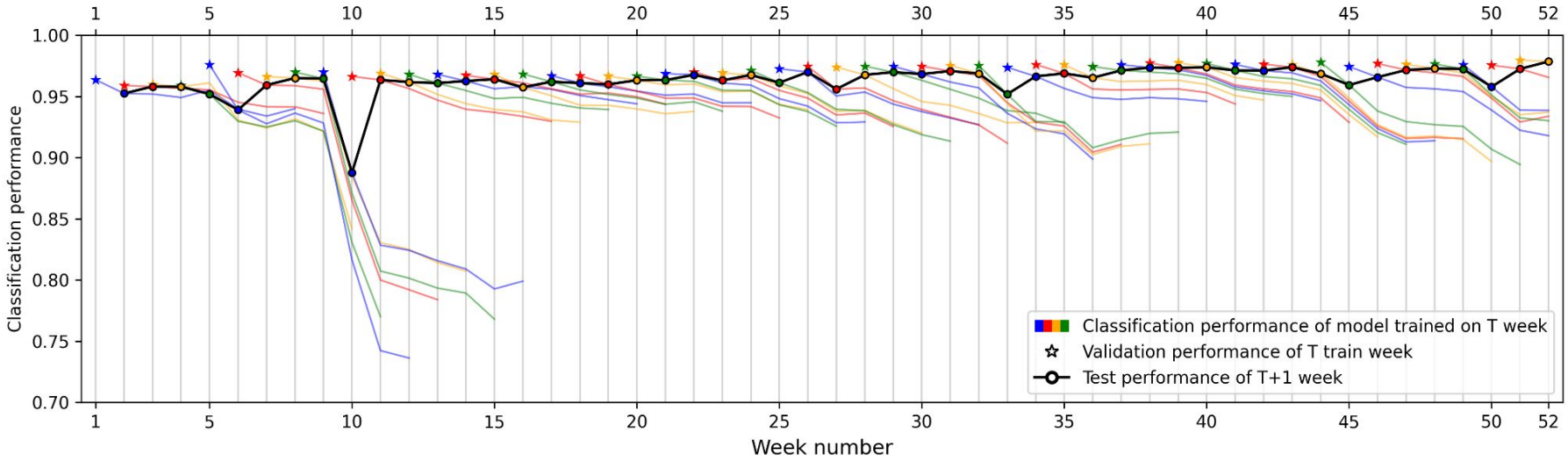
cesnet

Inference rate

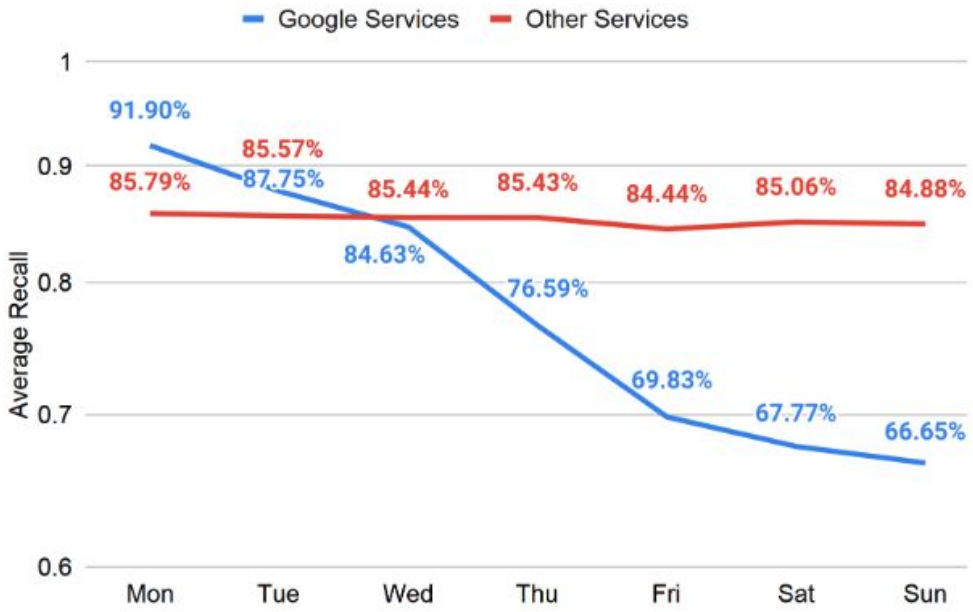
- ML model have 99.9% accuracy 💪
- CESNET network generate 300 000 flows/s
- 300 000 * (1 - 0.999) = 300 false classifications



Traffic Evolution



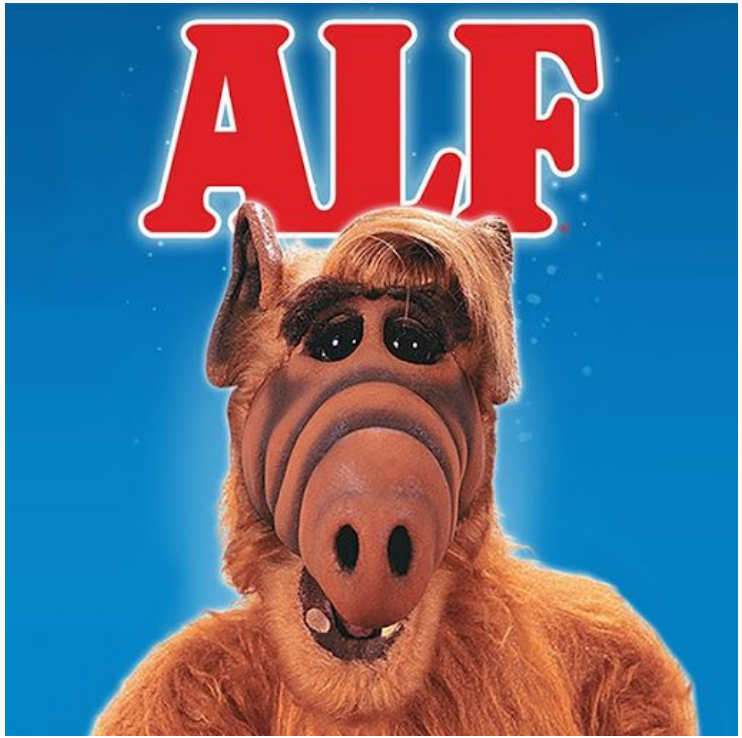
Traffic Evolution



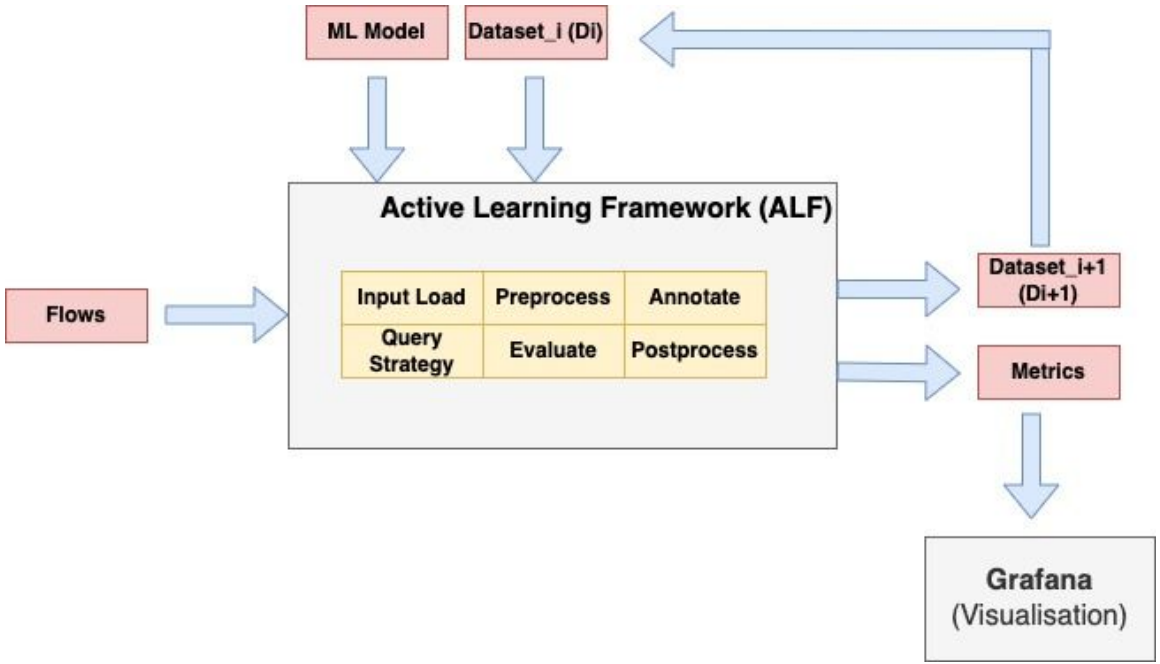
Data Drifts

- The inevitable change in the training data
- Network data are specific
  - Changes can be fast
  - Unpredictable

- Always working with trade-off ⚖️
  - Longevity
  - Accuracy


ALF

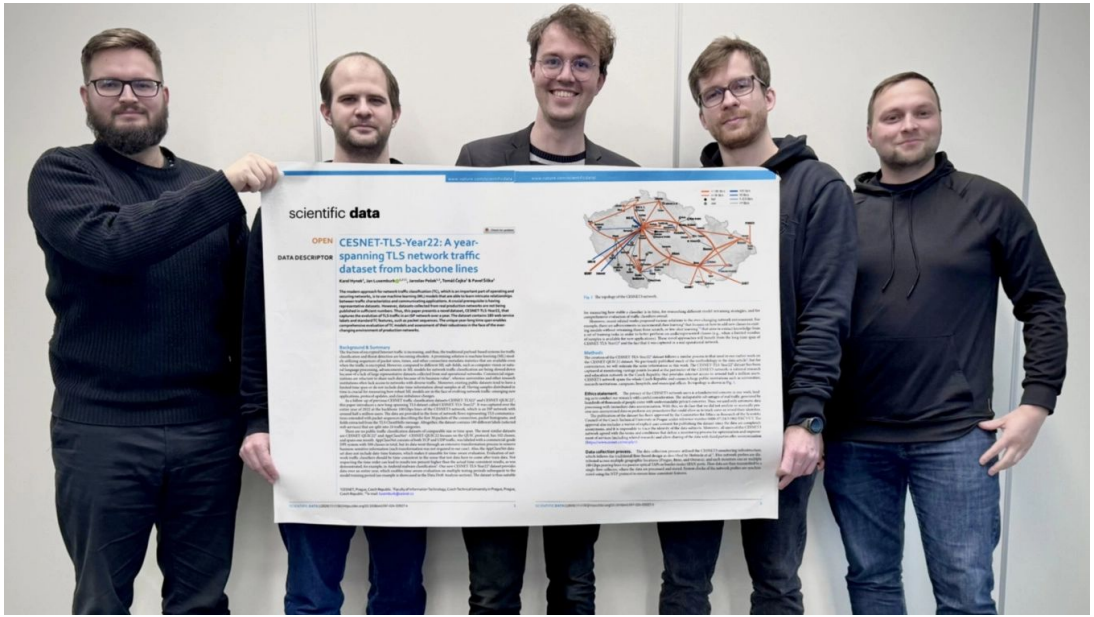
Active Learning Framework



- Iteratively re-train models
- Challenges in labeling
  - The active learning pipeline is selecting most important samples that should be labeled
- Cryptomining detection is being tested within this setup

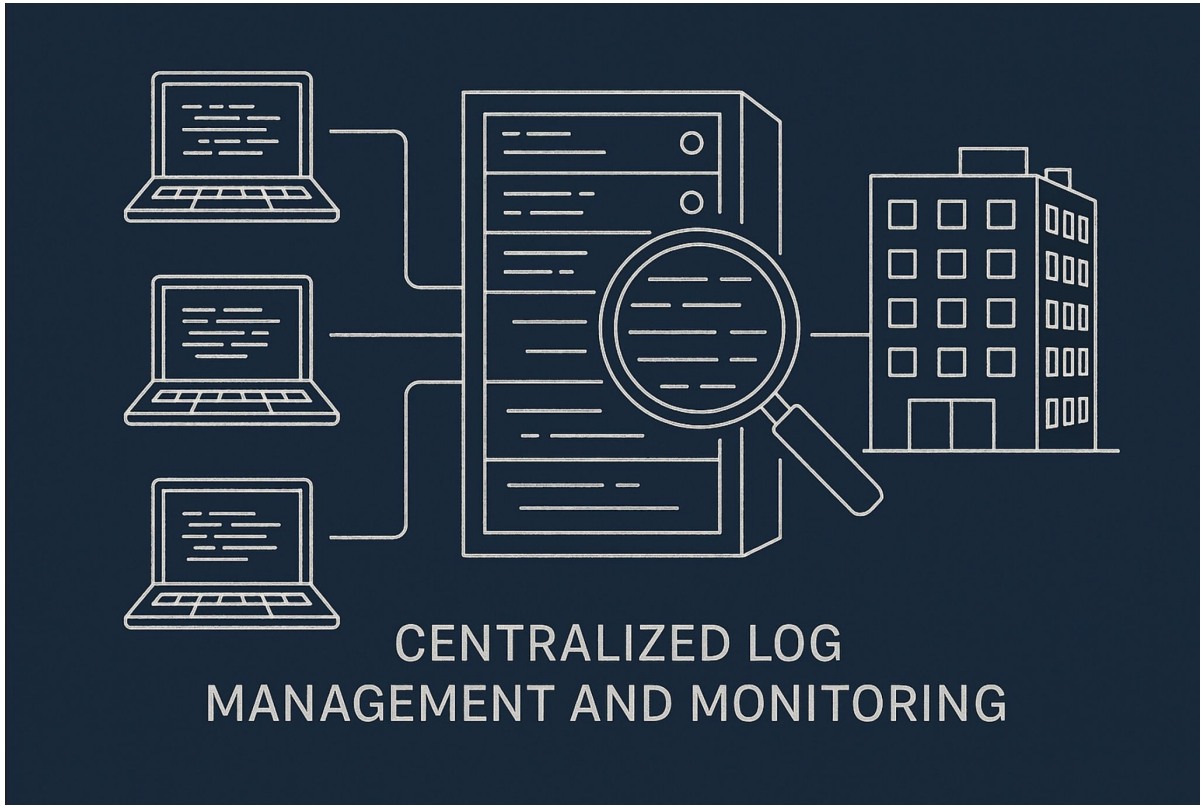
Temporal datasets

CESNET-TLS-Year22 – Whole 2022 Year of anonymized TLS traffic from CESNET network:
https://www.nature.com/articles/s41597-024-03927-4



cesnet

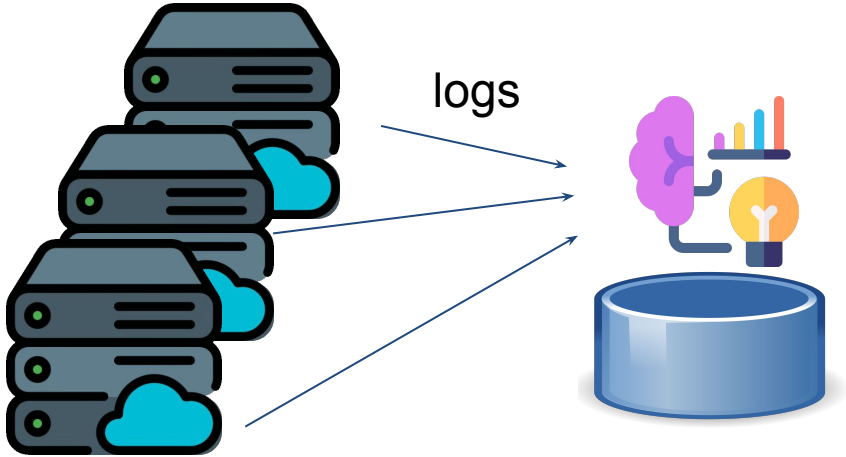
Logmanagement



- Motivation - central logging service at CESNET
- Commercial partner - Logmanager
- Analysis of top user issues during log management
  - customization to the logs of the particular infrastructure
  - custom alerts and dashboards
  - anomalies

Logmanagement - custom deployment

- Automatically recognize what is the source service?
- Imagine an infrastructure
  - with hundreds of services
  - new services appearing constantly
  - Tagging based on service
  - Optimized parsing



- Generating new parsers on not recognized logs

cesnet

Summary

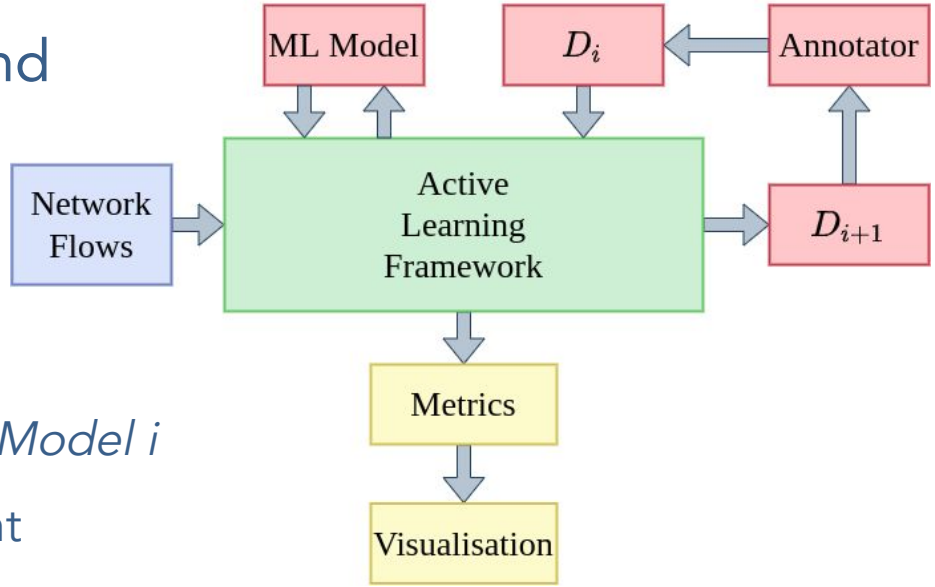
- Lessons learned
  - compare complex methods with basic ones (1-NN works well on flow data)
  - automate or avoid the need for annotated datasets
  - even 0.01% false positives may not be enough with high inference rate
  - assistant rather than the decision maker is preferred in production
- Publicly available datasets
- Transfer the results into production
- Future projects - AI for helpdesk, AI agent for SOC

cesnet

Thank you for your attention. Questions?

Annotation of datasets for encrypted traffic classification

- Flow-based encrypted traffic analysis using extended features/statistics and machine learning - when there is no way around



- Active Learning Framework (ALF)
  - continuous updates of *Dataset i* and ML *Model i*
  - research in Quality of Dataset Assessment
  - Goal: Automate and Keep up-to-date

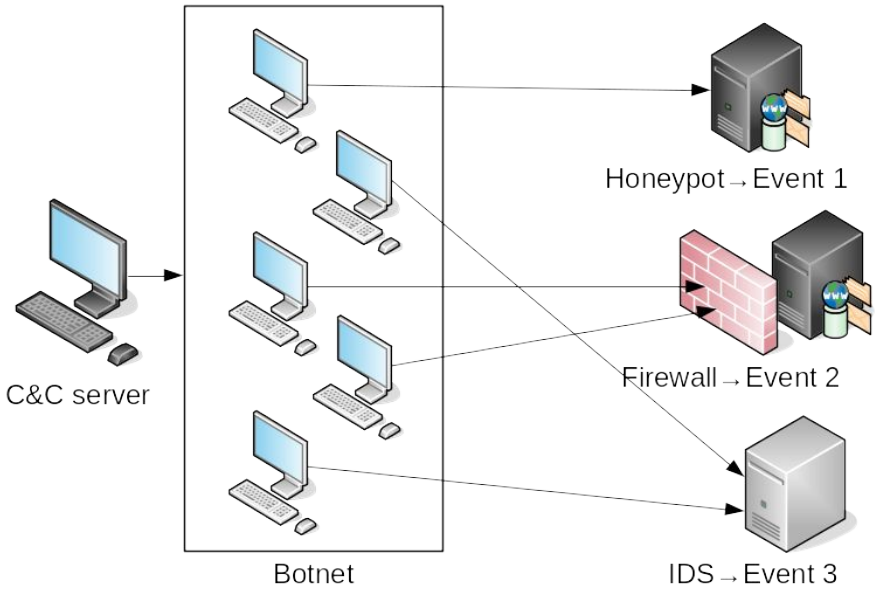
cesnet

Encrypted traffic datasets

- Characteristics
  - per-packet information (sizes, timestamps, directions, TCP flags)
  - packet histograms
  - packet bursts
  - …
- Results - HTTPS brute-forcing (FPR 0,0001)
- Lessons learned
  - different classifiers work with different features,
  - feature engineering can help to increase accuracy,
  - beware of false positives, it can be high number of “false alerts”,
  - quality of publicly available datasets is bad sometimes

Discovering coordinated IP addresses

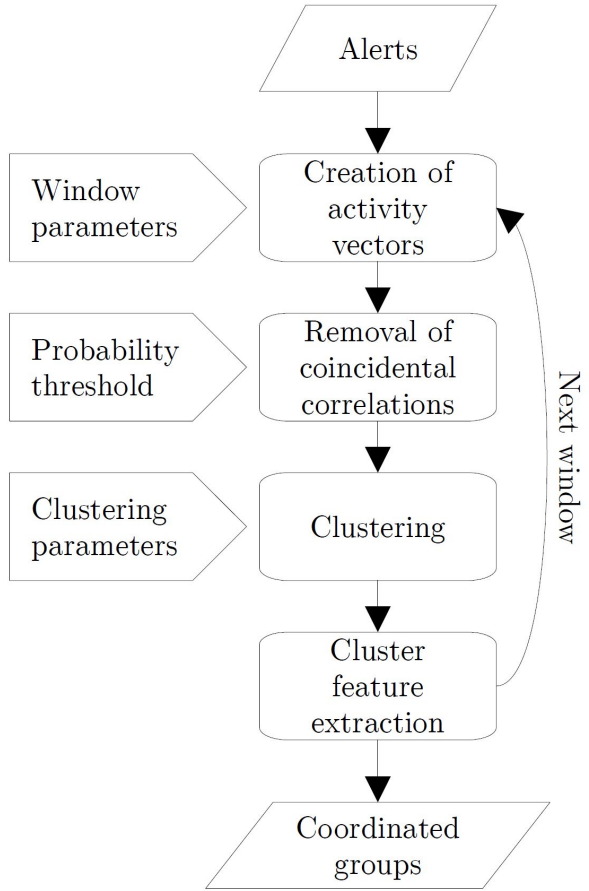
- Suspicious coordinated behavior is an indication of botnet activity
- Our goal is to discover groups of IP addresses that exhibit temporal correlation



Discovering coordinated IP addresses

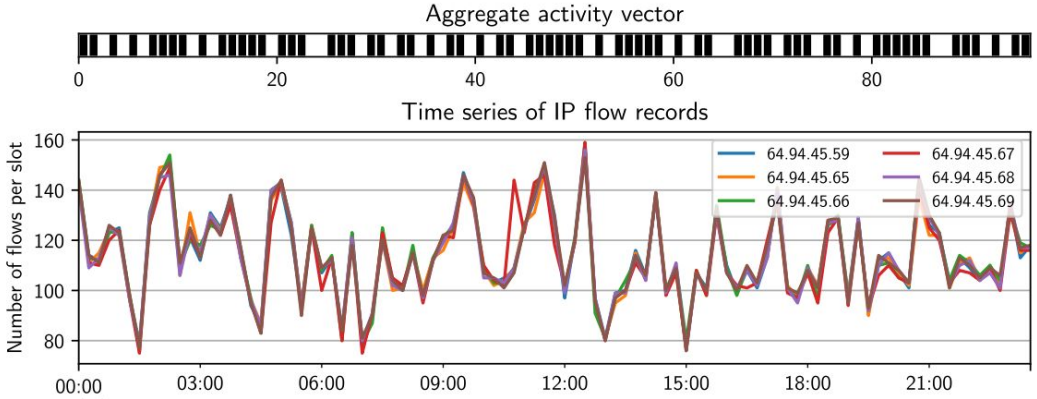
- Vectors of activity per each IP
- Unsupervised ML

Slot/ Entiy	1	2	3	4	5	...	n-2	n-1	n
1	•			•	•		•		•
2	•			•			•	•	
3	•	•	•	•	•			•	•
...									
m		•	•	•			•		•



Discovering coordinated IP addresses

- Results



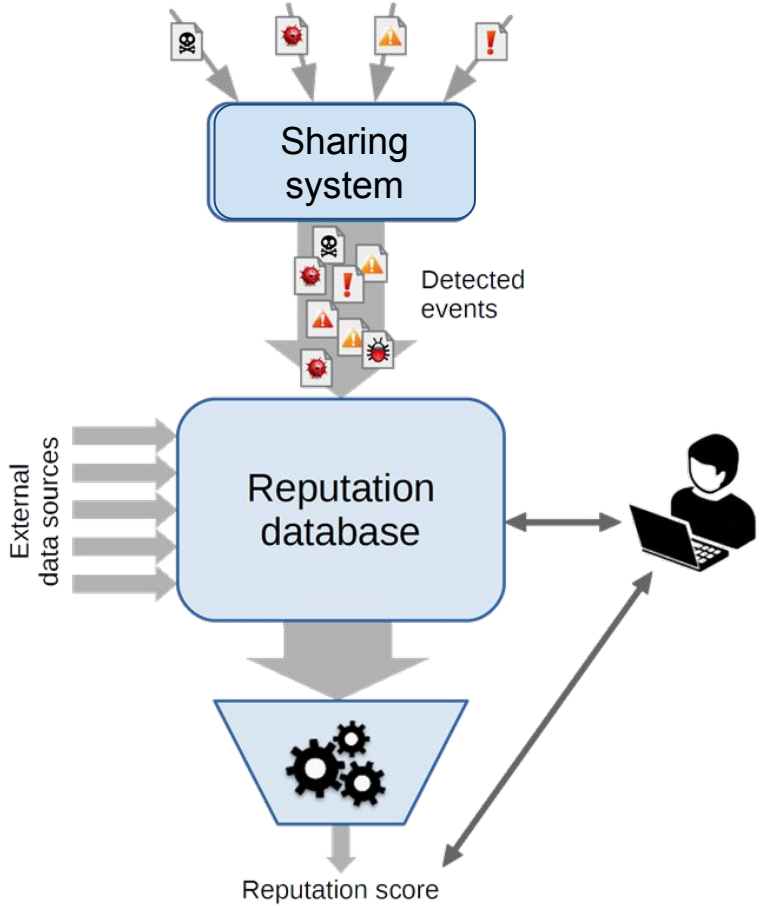
- Lessons learned
  - Verification of clusters is crucial not to detect legitimate coordination
  - Accidental coordination due to high data volume
  - Preprocessing and clustering is the key otherwise vectors in the clusters can be very different to each other
  - Bursty behavior vs. periodic vs. random
  - Bad neighbourhood discovery

Blocklists based on reputation

- Learning from what we observed is useful to be prepared in the future
- Predictively blocking IP addresses. Ok. But which one?
- Network Entity Reputation Database (NERD)

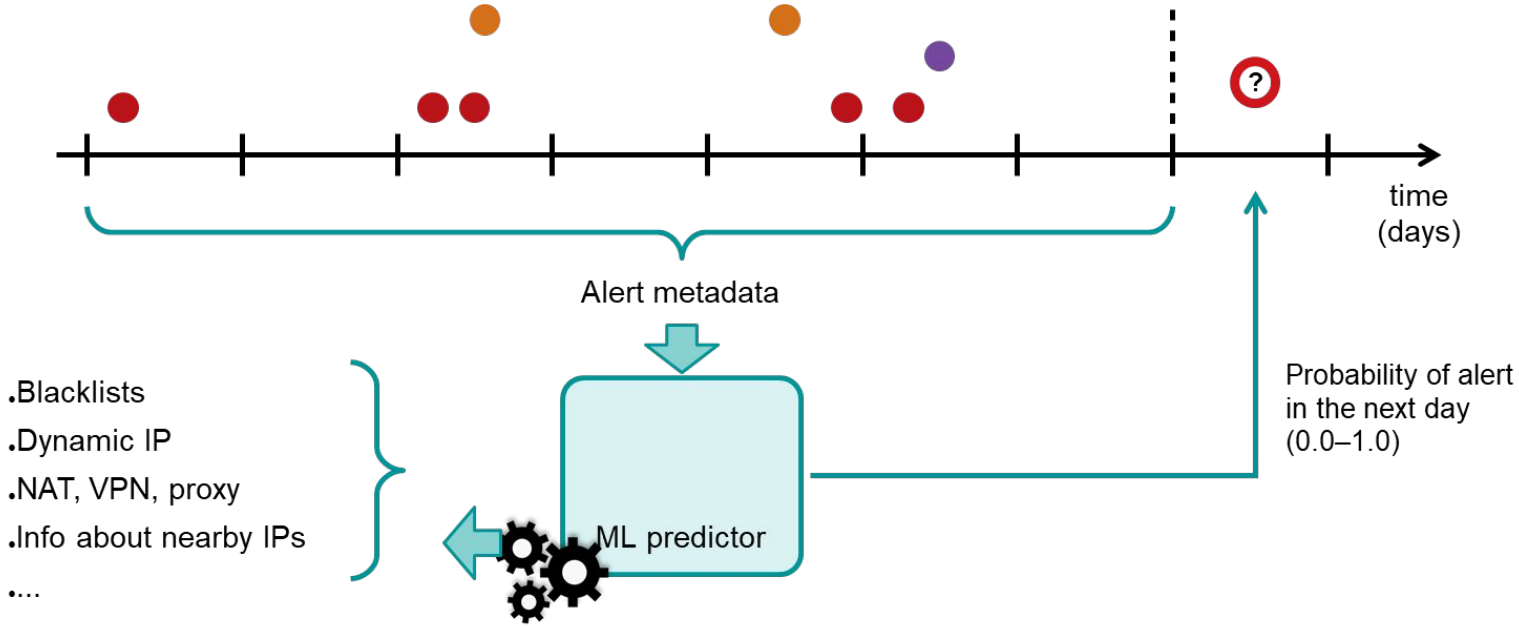
Blocklists based on reputation

- Collected data from IDS, honeypots, NBA, …
- Enhanced by contextual data from external data sources
  - DNS, whois, geolocation
  - Blacklists, …
  - Shodan, VirusTotal
  - Openresolvers
  - TOR exit nodes



Blocklists based on reputation

- ML calculates reputation of an IP address
- The reputation predictor learns itself based on history



cesnet

Inference of DDoS mitigation rules

Results

- max 3% false positives, more than 98% true positives
- tested on multiple multi-vector attack samples
- integrated in CESNET DDoS Protector

Lessons learned

- No annotated dataset needed, training takes place during DDoS attack
- Counterintuitive hyperparameter tuning
- Works very well as long the majority of traffic is DDoS
- Needs performance optimization for real deployment
- Streaming variant forgets very fast what is normal

Blocklists based on reputation

- Results
  - Take top-N IP address according to reputation score and it creates the most effective blocklist of a given size
- Lessons learned
  - Other use cases: alert prioritization, quick overview for an analyst
  - Different types of attacks can be predicted separately but data needed to achieve accuracy (100,000s alerts of each type)
  - Model performance is quite stable even after a few months
  - Model can be easily extended to use data from new sources

Summary

- Use AI
  - if data annotation is cheap (can be automated or avoided)
  - if there is no easier way
- High quality data (bias, noise, correlations)
- Good data preprocessing
- Use AI method to fit your need
  - ML is not the only AI option (e.g. information fusion)
- Even good results may not be sufficient (low false positive rate translates to many false positives in absolute numbers)