cesnet

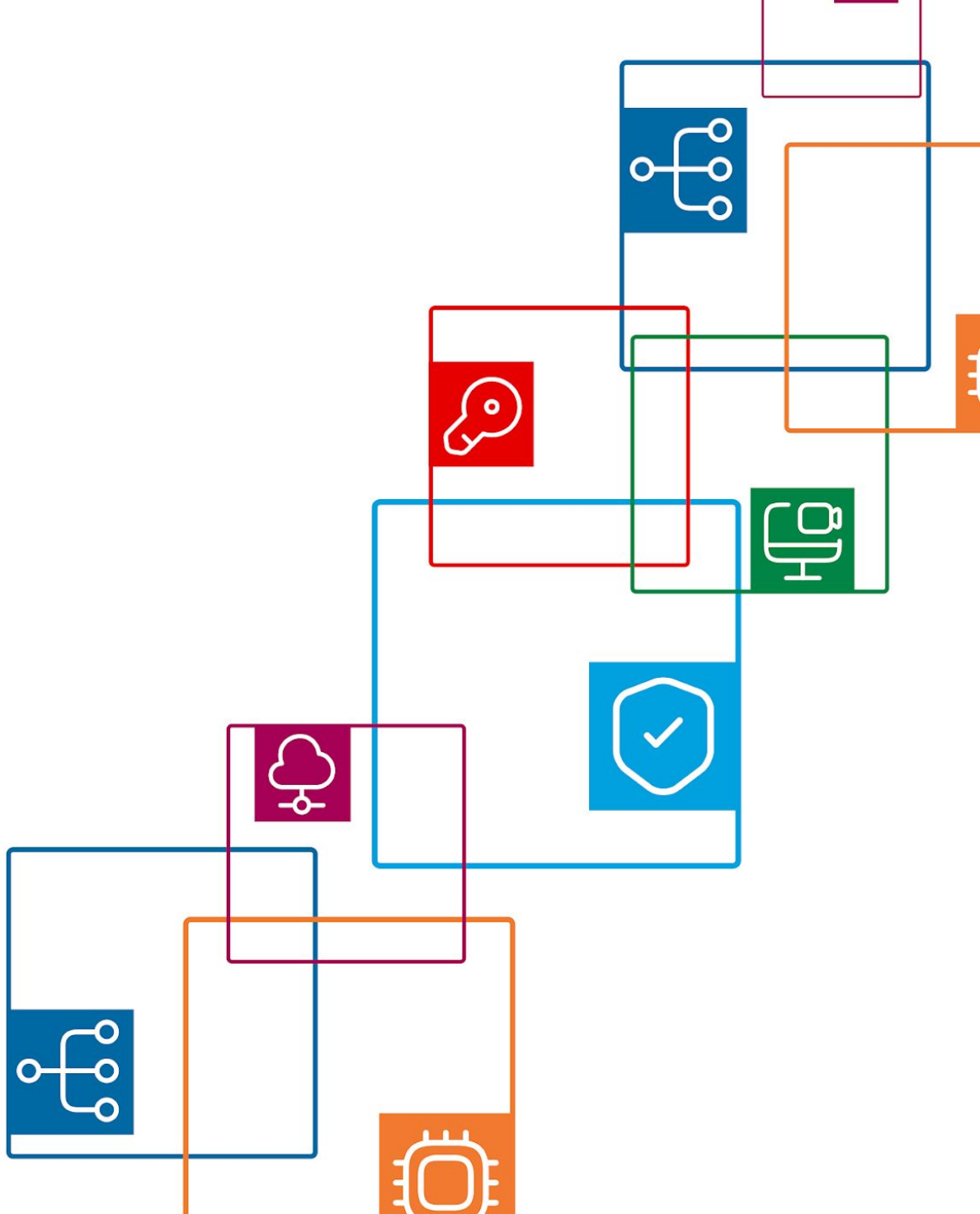
SUBMERSE

Tour into Practical AI for Network Traffic Analysis

Jaroslav Pesek

Department of Administration and Security Tools

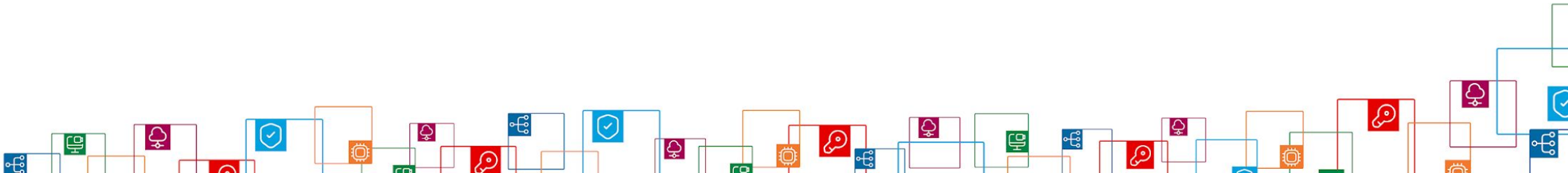
pesek@cesnet.cz



cesnet

Traffic Analysis Team

- One part of the department:
  - 2 post-docs
  - 8 PhD students
  - 3 full time developers
  - uncountable number of students (~ 10)
- Main topics:
  - Development of tools and software
  - Dataset creation
  - Research of application of machine learning
  - Building of infrastructure for handling the data

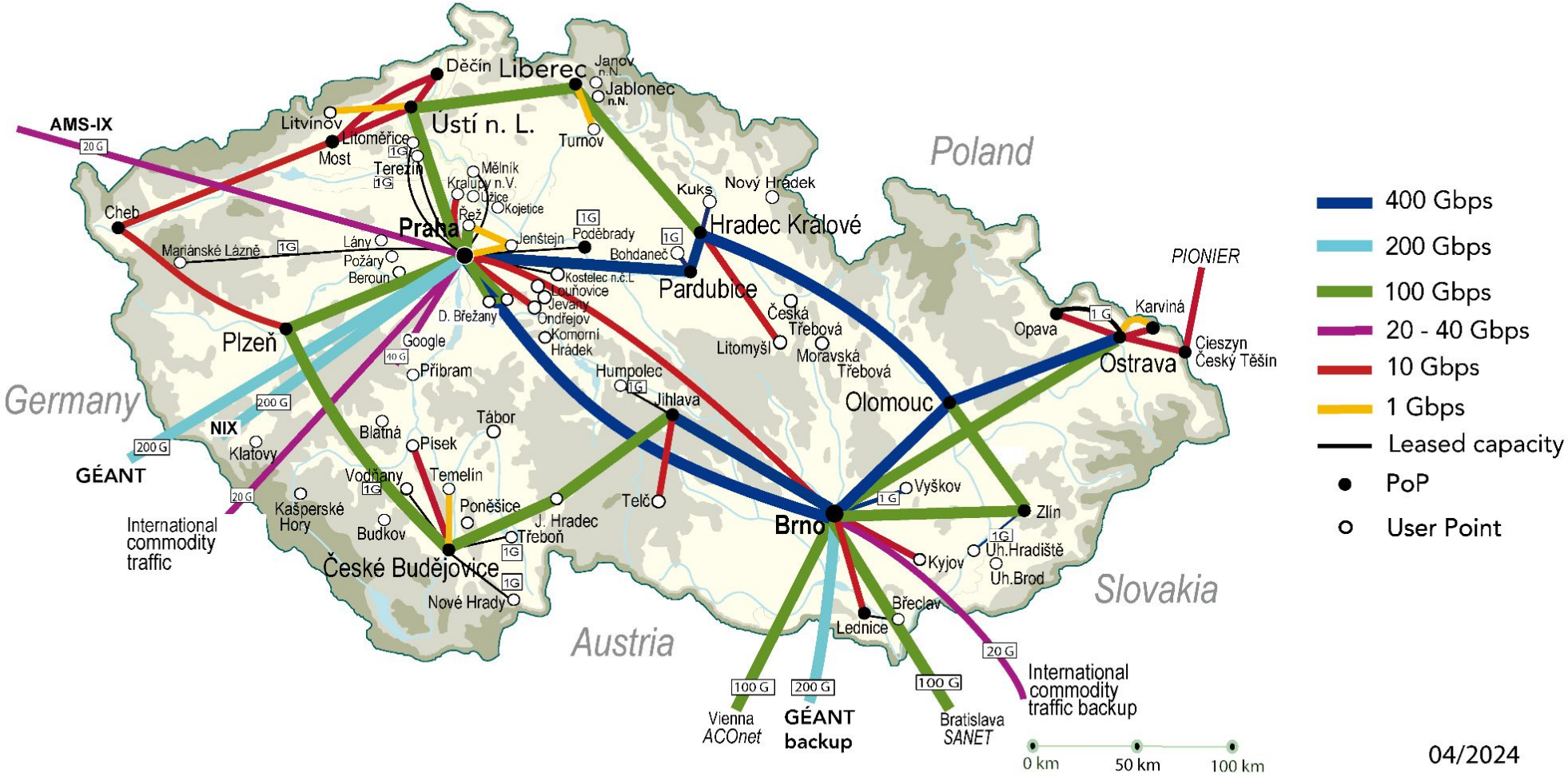


cesnet

Outline

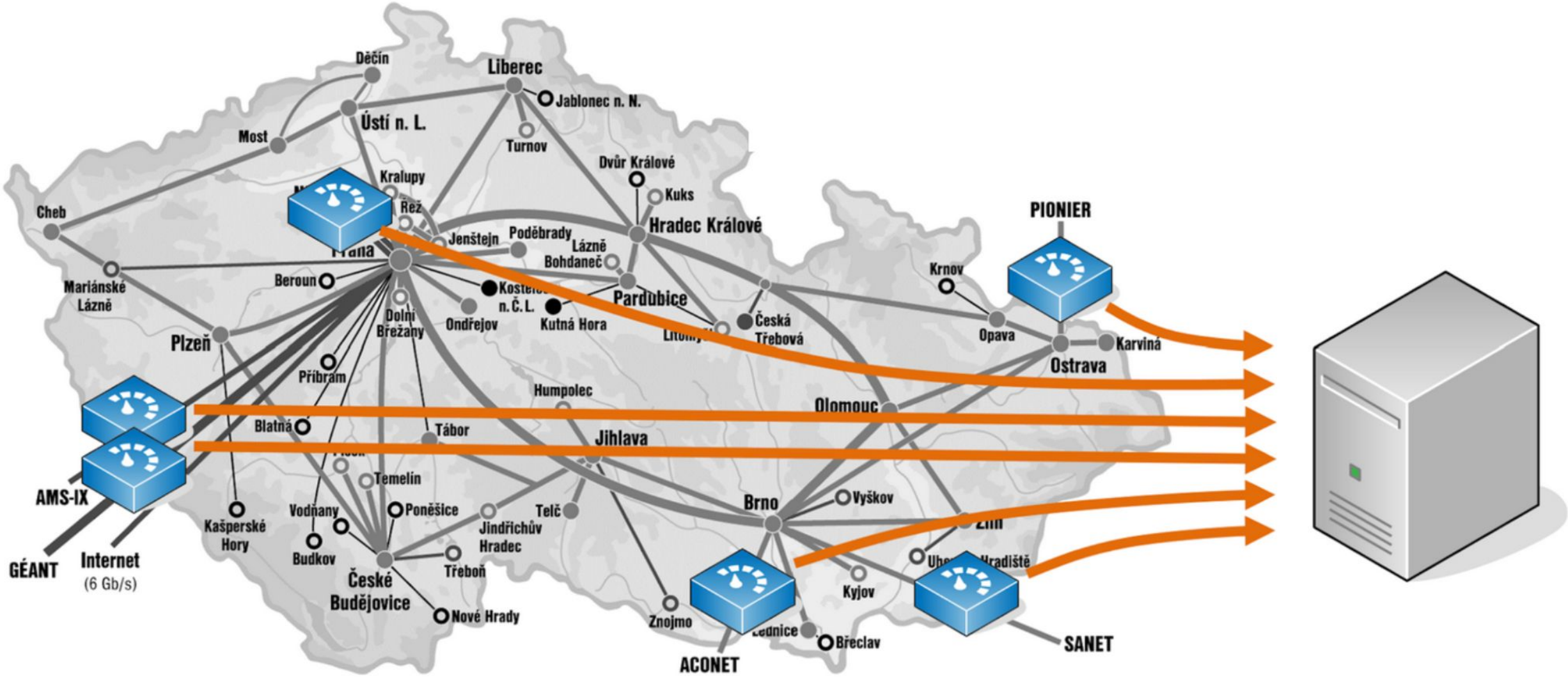
1. Infrastructure and software, visualization and LLM baby steps
2. Dataset and tools
3. Is field in crisis?

Network

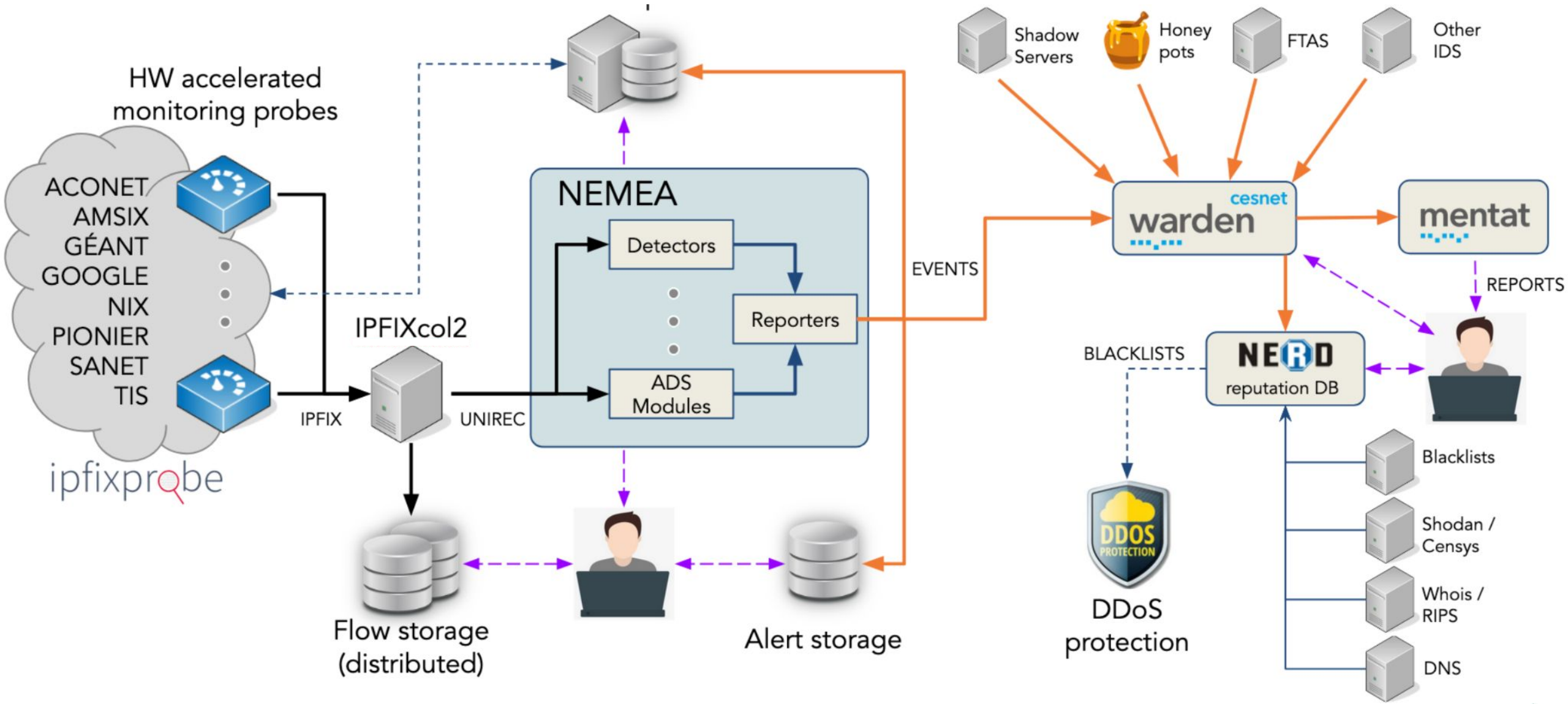


04/2024

Probes


Děčín
Liberec
Jablonec n. N.
Ústí n. L.
Most
Turnov
Dvůr Králové
Kralupy
Kuks
Cheb
Hradec Králové
PIONIER
Jenštejn
Poděbrady
Lázně Bohdaneč
Mariánské Lázně
Beroun
Pardubice
Krnov
Dolní Břežany
Ondřejov
Kutná Hora
Česká Třebová
Opava
Karviná
Plzeň
Ostrava
Příbram
Humpolec
Olomouc
Blatná
Tábor
Jihlava
AMS-IX
Temelín
Vyškov
Kašperské Hory
Vodňany
Poněšice
Telč
Brno
Budkov
Jindřichův Hradec
Třeboň
České Budějovice
Kyjov
Znojmo
Nové Hrady
Břeclav
GÉANT
Internet
(6 Gb/s)
ACONET
SANET

Traffic monitoring and analysis



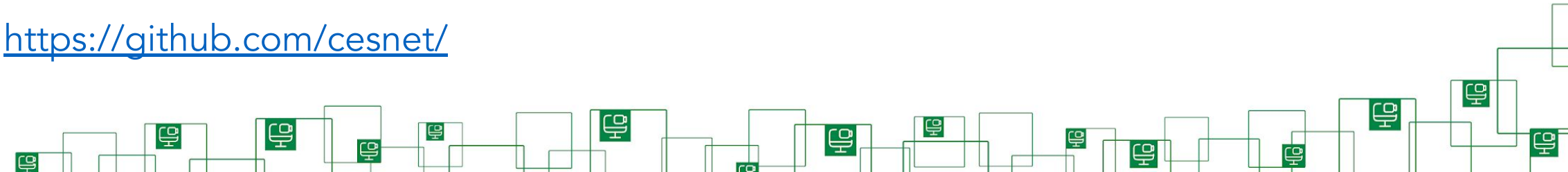
cesnet

Tools and software

Production-ready, established tools

- ipfixprobe – modular and powerful flow exporter
- IPFIXcol2 – flow collector, supports IPFIX, NetFlow v5/v9
- NEMEA – modules based framework for data stream analysis

https://github.com/cesnet/

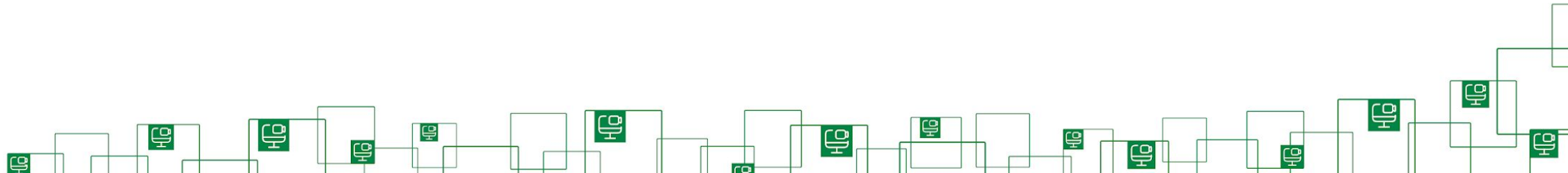


cesnet

Degustation of our tech stack

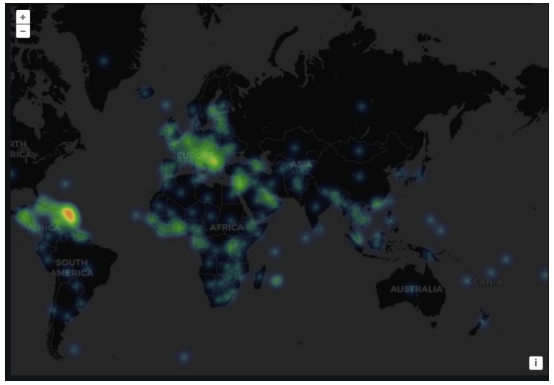
Besides our in-house software, we use

- Nix and NixOS for reproducible and fast deploying (we replaced Ansible)
- Clickhouse as analytical database
- Grafana for visualization



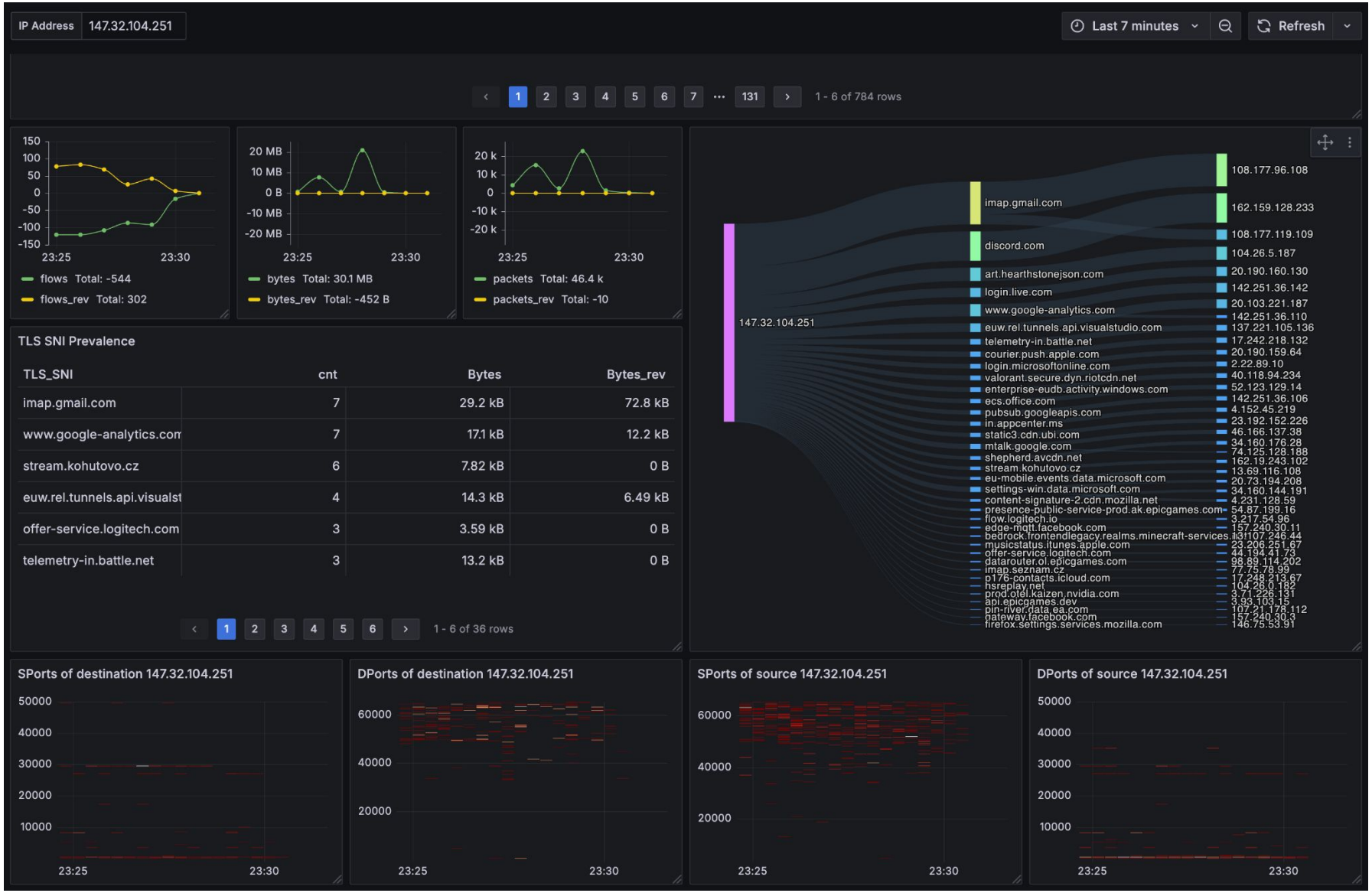
Degustation of our tech stack

cesnet



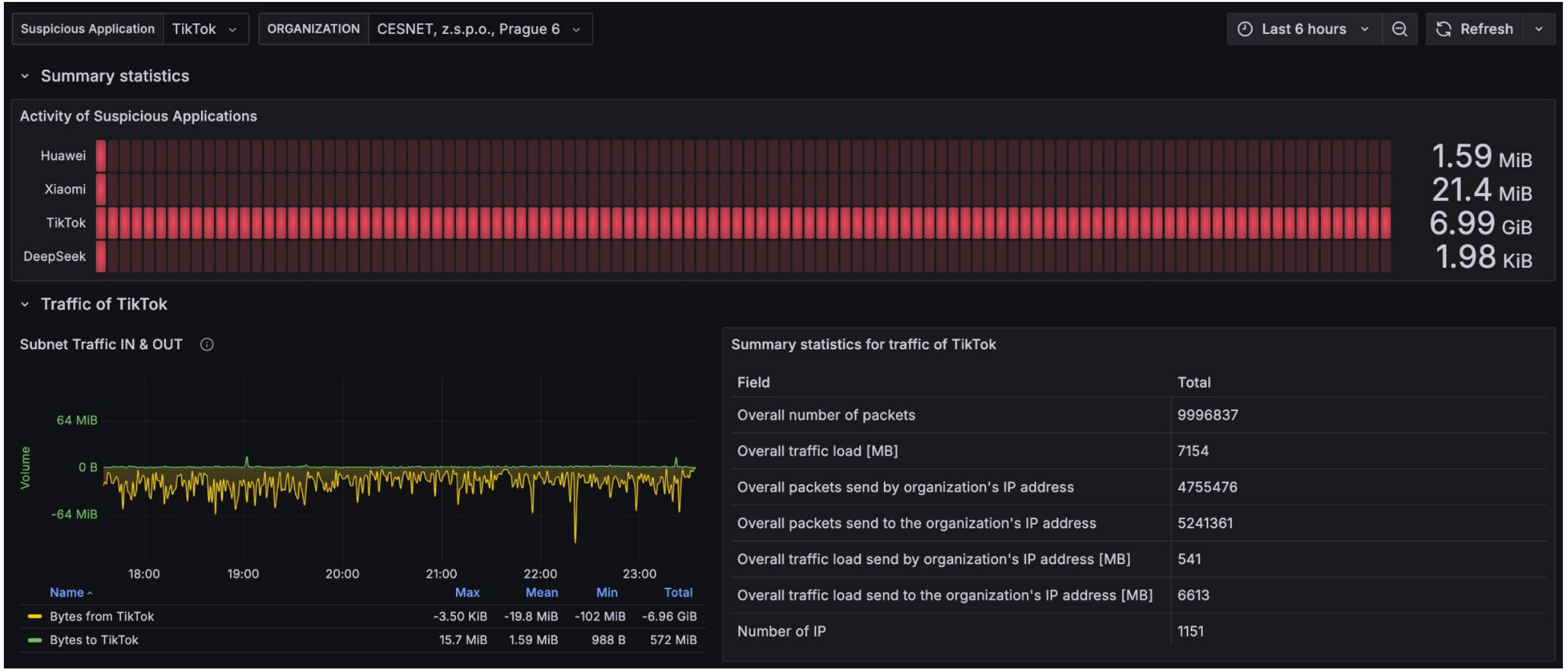
Degustation of our tech stack

cesnet



Degustation of our tech stack

cesnet



LLM Baby steps

We try to use LLM agent that reads the dashboards and reports to superior, i.e.:

Daily SOC Report – 2025-10-08

Network activity remained within normal limits. No major anomalies or security events detected overnight.

Inbound and outbound traffic volumes were balanced (≈55 % / 43 %), dominated by TCP (≈90 %) with minor UDP activity. Top talkers and ports matched expected operational behavior.

The “Decrypto” detection module reported no critical findings or suspicious flows. Geolocation map shows usual communication patterns—mainly Western Europe and North America—with no unexpected hotspots.

Action: No escalation required. Continue regular monitoring and verify top hosts if traffic growth persists.

LLM Baby steps

cesnet

Or:

Daily SOC Report – 2025-10-08

A volumetric anomaly was detected between 02:15–03:00 UTC, indicating a possible DDoS attack targeting internal web services. Inbound traffic spiked to over **8× the daily average**, primarily consisting of **UDP** and **TCP SYN** packets.

Geolocation analysis shows that over **70 % of the malicious traffic** originated from IP ranges registered in Russia, with smaller portions from Ukraine and Belarus. The top destination was **147.32.1.10 (public web endpoint)**, receiving sustained requests on ports **80** and **443**.

Firewall counters confirmed **packet drops and connection resets** consistent with rate-limiting behavior. No data exfiltration or lateral movement observed.

Action: Incident escalated to L2 for correlation and mitigation. Recommend reviewing upstream filtering rules and updating threat intelligence lists.

LLM Baby steps

LLM as SOC member:

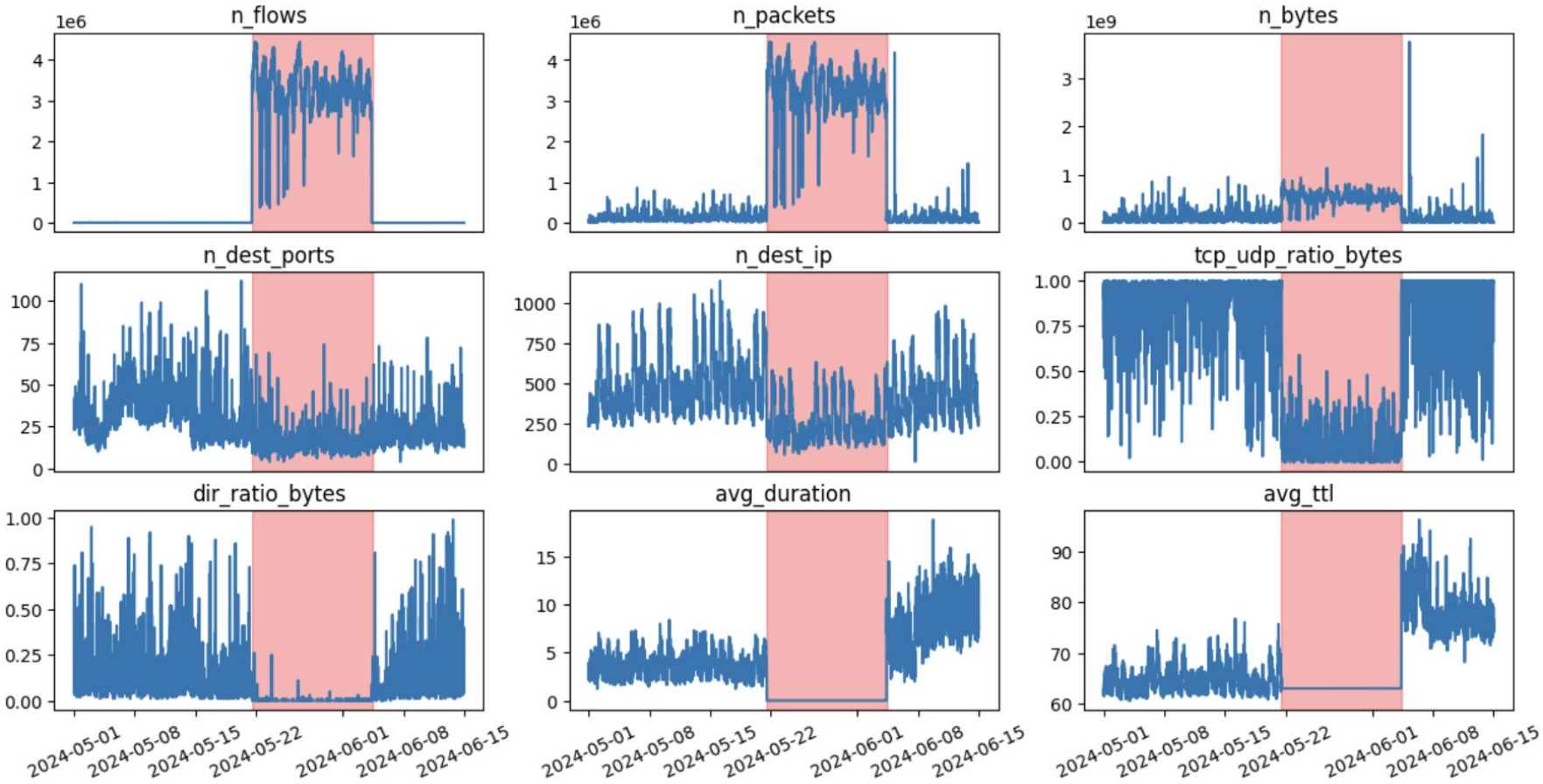
Label: Potential DDoS

Priority: critical

Confidence: High

Explanation: Significant spikes in multiple metrics suggest a coordinated traffic anomaly.

Suggested Action: Immediately investigate the source IPs and traffic patterns; implement rate limiting if confirmed as an attack.



cesnet

Dataset tools

CESNET-DataZoo & CESNET-TsZoo

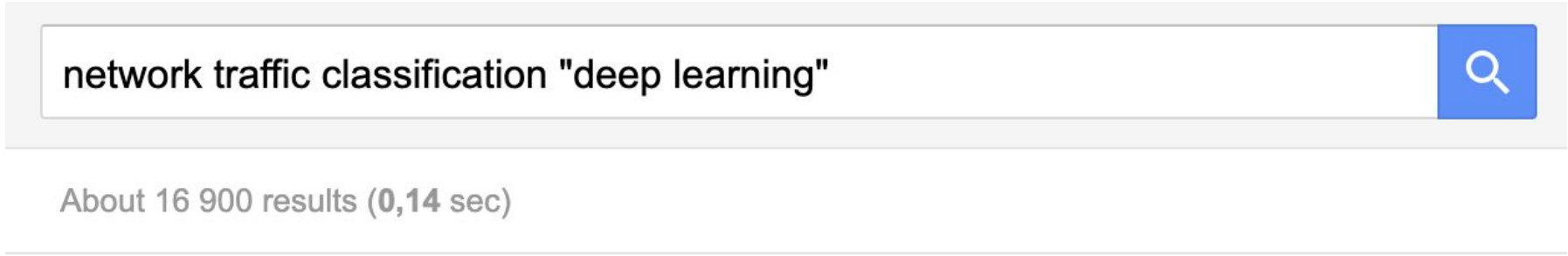
- Tools for working with large REAL traffic data in simple way
- CESNET-TLS-Year22 and CESNET-TimeSeries24 published in Nature Scientific Data journal
- QUIC services dataset
- WIP: Curating new benchmark datasets from outside

```python
from cesnet_datazoo.datasets import CESNET_QUIC22
from cesnet_datazoo.config import DatasetConfig, AppSelection

dataset = CESNET_QUIC22("/datasets/CESNET-QUIC22/", size="XS")
dataset_config = DatasetConfig(
    dataset=dataset,
    apps_selection=AppSelection.ALL_KNOWN,
    train_period_name="W-2022-44",
    test_period_name="W-2022-45",
)
dataset.set_dataset_config_and_initialize(dataset_config)
train_dataframe = dataset.get_train_df()
val_dataframe = dataset.get_val_df()
test_dataframe = dataset.get_test_df()
```

When simple model is just enough

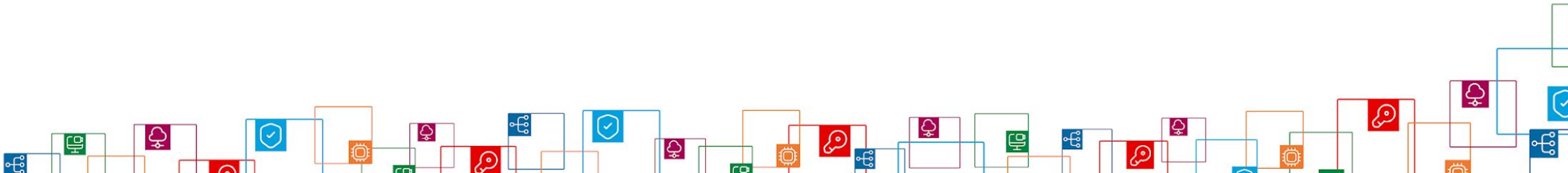
- Contemporary research often uses complex, deep learning models
- Papers published in 2025 only…:



cesnet

When simple model is just enough

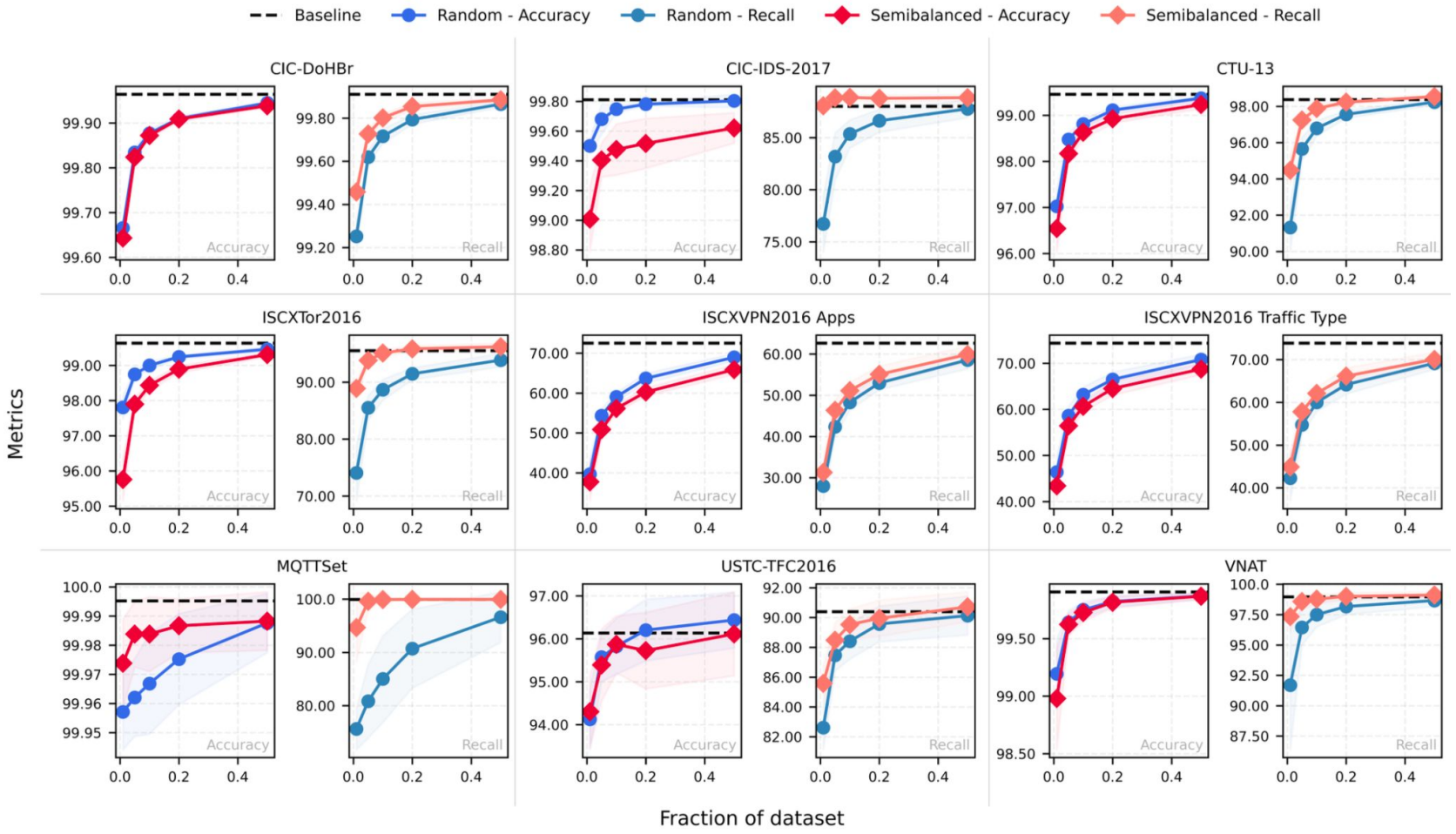
- Contemporary research often uses complex, deep learning models
- Our study shows simple 1-NN (one nearest neighbour) using sequence of 10 first packets **only** can achieve the similar performance
- The model is not genius, it is data redundancy in benchmark datasets
- The reported progress in field is overestimated



Findings

Even when we heavily sample the dataset, the results are still quite ok

We see the same phenomenon for all major community favorit datasets

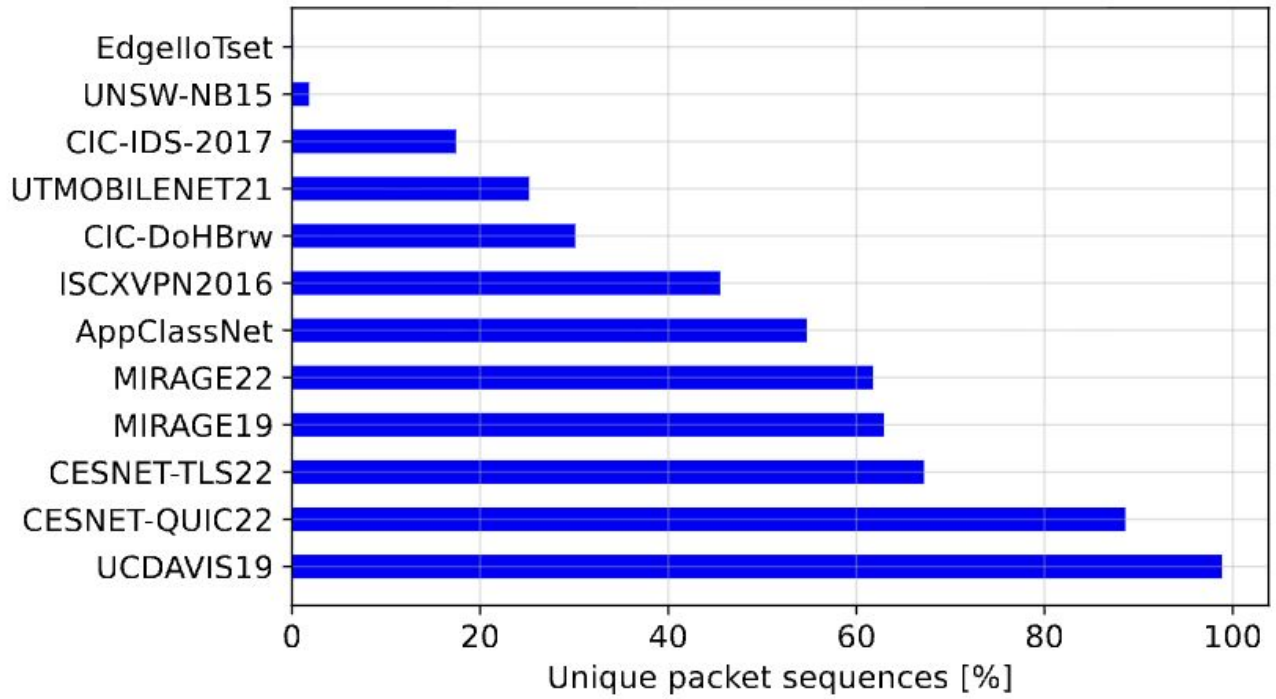


Findings

Fraction of redundant sequences are relatively high

Models just memorize samples

That is why 1-NN model just works



Comparison

cesnet

Dataset	Task	SOTA Accuracy [%]	Baseline Accuracy [%]	Δ Accuracy [pp]
ISCXTor2016 [10]	tor, non-tor	100.00 ± 0.01 [11]	99.63 ± 0.02	-0.37 ± 0.02
USTC-TFC2016 [12]	10 mal., 1 benign	98.30 [13]	96.14 ± 0.64	-2.16 ± 0.64
VNAT [14]	vpn, non-vpn	98.03 [15]	99.91 ± 0.01	1.88 ± 0.01
CTU-13 [3]	mal., benign	99.30 [16]	99.46 ± 0.08	0.16 ± 0.08
MQTTSet [17]	4 mal., 1 benign	99.90 [18]	100.00 ± 0.01	0.10 ± 0.01
ISCXVPN2016 App [1]	17 benign svc.	79.92 ± 1.31 [19]	72.54 ± 1.05	-7.38 ± 1.68
ISCXVPN2016 Traffic [1]	6 benign types	81.71 ± 1.26 [19]	74.39 ± 1.35	-7.32 ± 1.85
CIC-DoHBr [20]	mal., benign	99.99 [21]	99.96 ± 0.00	-0.03 ± 0.00
CIC-IDS-2017 [2]	7 mal., 1 benign	95.79 [22]	99.81 ± 0.02	4.02 ± 0.02

cesnet

Takeaways

- We should focus less on building bigger models and more on building better, more diverse datasets that truly reflect the complexity of real network traffic
- Let's create redundancy-aware benchmarks
- Is traffic analysis in crisis?

Based on papers:

https://ieeexplore.ieee.org/abstract/document/11096965

https://arxiv.org/pdf/2506.08655