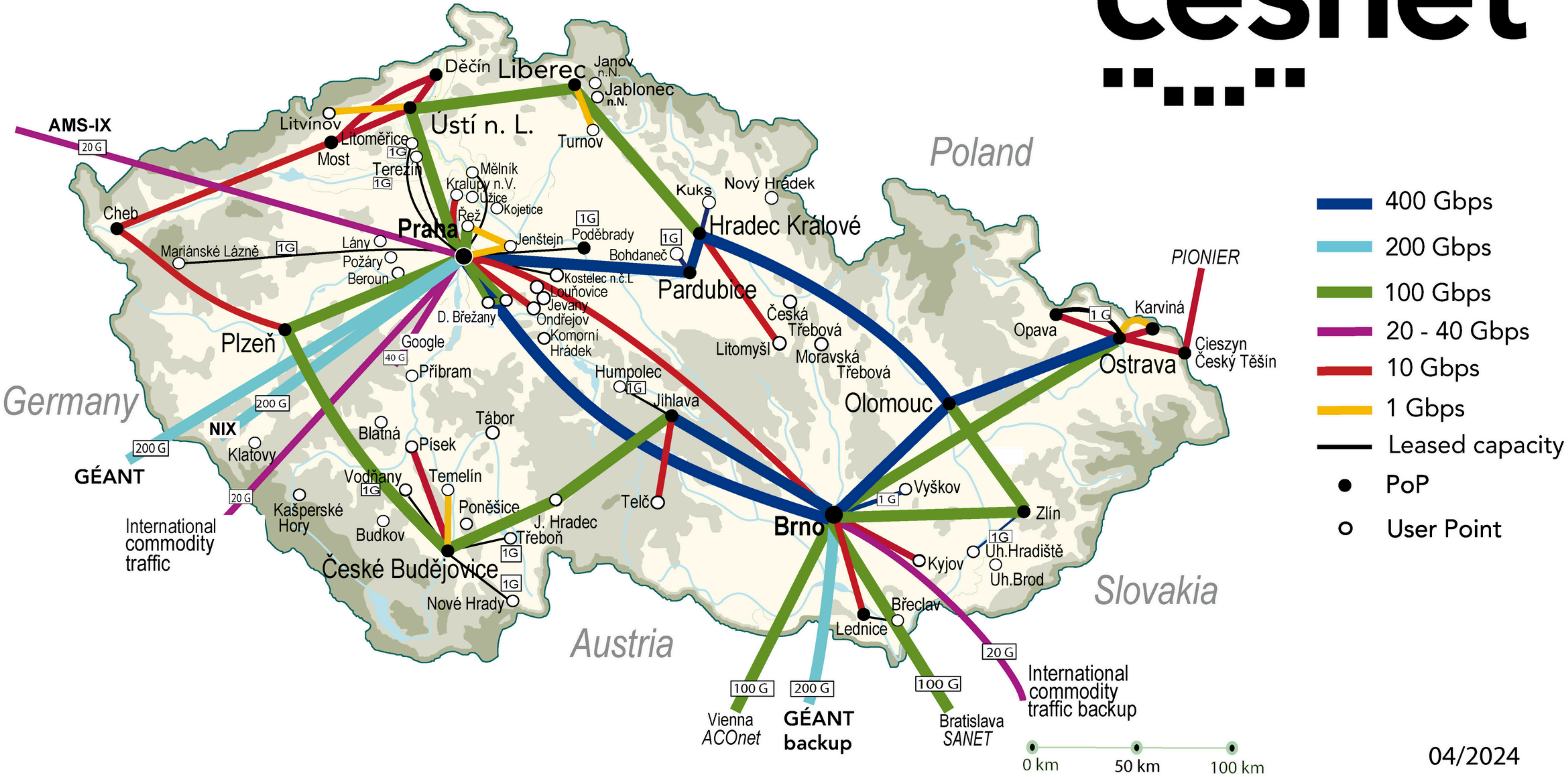
# ANOMALY DETECTION IN ISP NETWORKS

***Josef Koumar**, CESNET a.l.e. & CTU in Prague*

# MOTIVATION

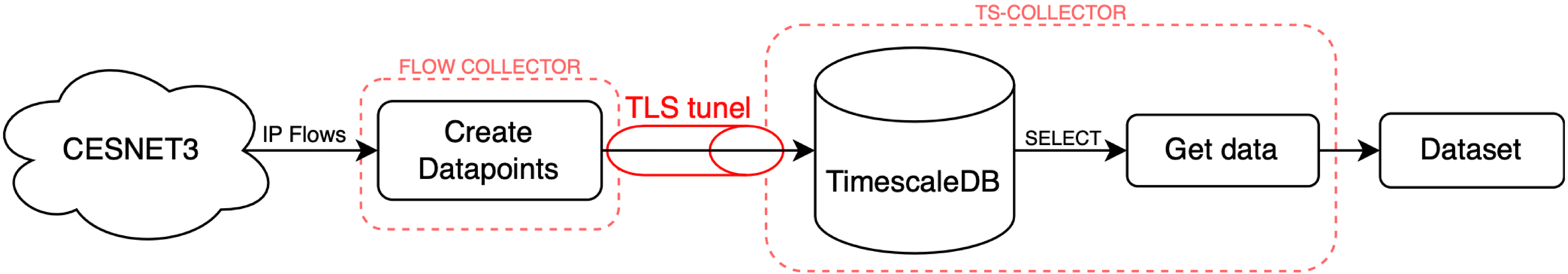
- There is lack of a reference datasets for network traffic forecasting and anomaly detection! --> Crucial obstacle identified in recent surveys.
- Additionally, real-world datasets used in the evaluation are not publicly available due to privacy concerns.

# DATASET CREATION

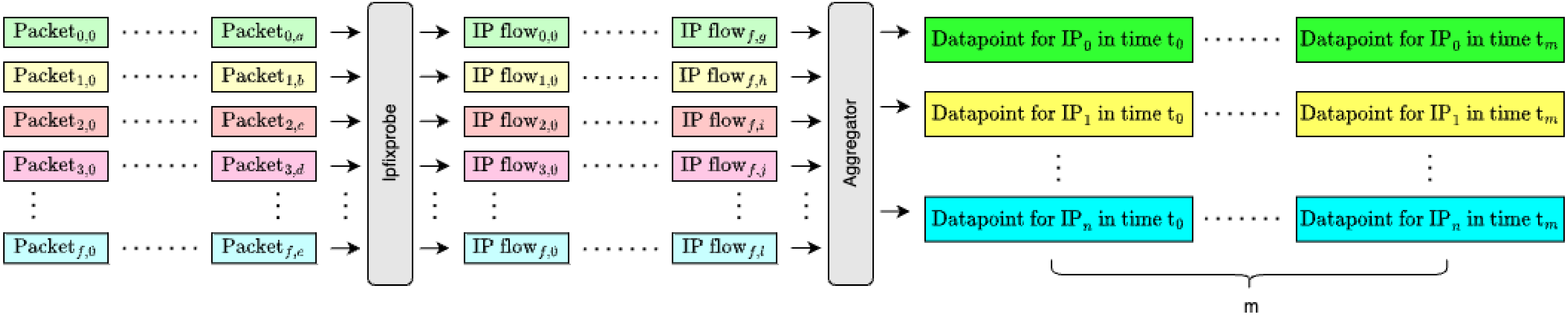


04/2024

# DATASET CREATION



# DATASET CREATION



$\text{Packet}_{0,0}$ ······· $\text{Packet}_{0,a}$
$\text{Packet}_{1,0}$ ······· $\text{Packet}_{1,b}$
$\text{Packet}_{2,0}$ ······· $\text{Packet}_{2,c}$
$\text{Packet}_{3,0}$ ······· $\text{Packet}_{3,d}$
$\text{Packet}_{f,0}$ ······· $\text{Packet}_{f,e}$
Ipfixprobe
$\text{IP flow}_{0,0}$ ······· $\text{IP flow}_{f,g}$
$\text{IP flow}_{1,0}$ ······· $\text{IP flow}_{f,h}$
$\text{IP flow}_{2,0}$ ······· $\text{IP flow}_{f,i}$
$\text{IP flow}_{3,0}$ ······· $\text{IP flow}_{f,j}$
$\text{IP flow}_{f,0}$ ······· $\text{IP flow}_{f,l}$
Aggregator
Datapoint for $\text{IP}_0$ in time $t_0$ ······· Datapoint for $\text{IP}_0$ in time $t_m$
Datapoint for $\text{IP}_1$ in time $t_0$ ······· Datapoint for $\text{IP}_1$ in time $t_m$
Datapoint for $\text{IP}_n$ in time $t_0$ ······· Datapoint for $\text{IP}_n$ in time $t_m$
m

# DATASET CREATION

Created from **66 billion IP flows** that contain **4 trillion packets** that carry approximately **3.7 petabytes of data**

Time Series Metrics:

- Number of IP flows, packets, bytes
- Number of unique destination IP addresses
- Number of unique destination ASNs
- Number of unique destination countries
- TCP/UDP ratio
- Packet direction ratio
- Average TTL and duration of IP flows

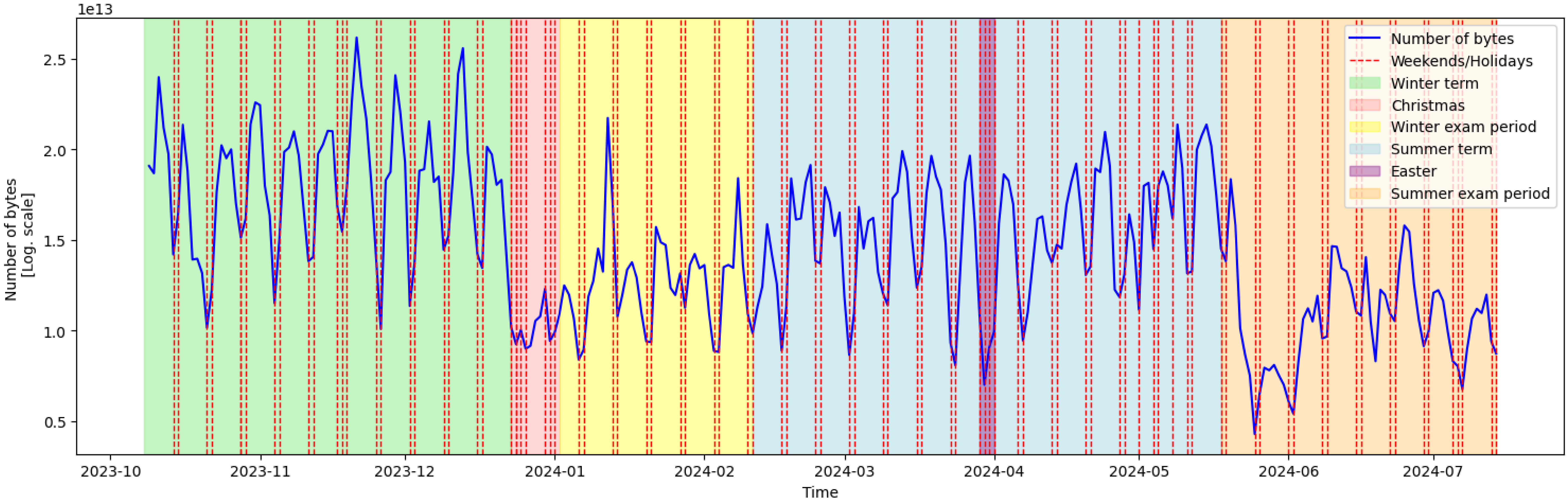
Aggregation:

- 10 minutes
- 1 hour
- 1 day

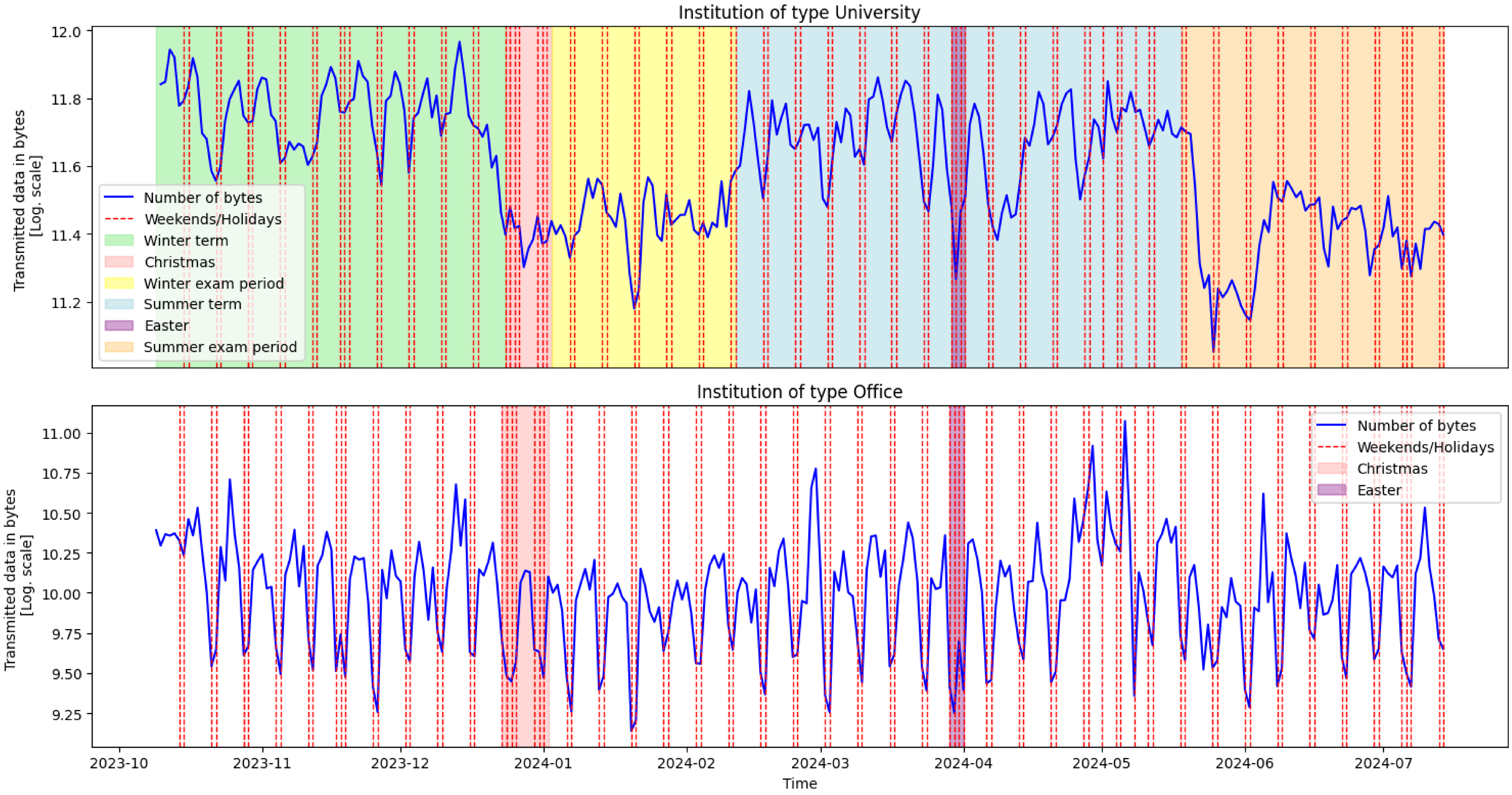
Identifiers:

- IP addresses
- Institutions
- Institution subnets

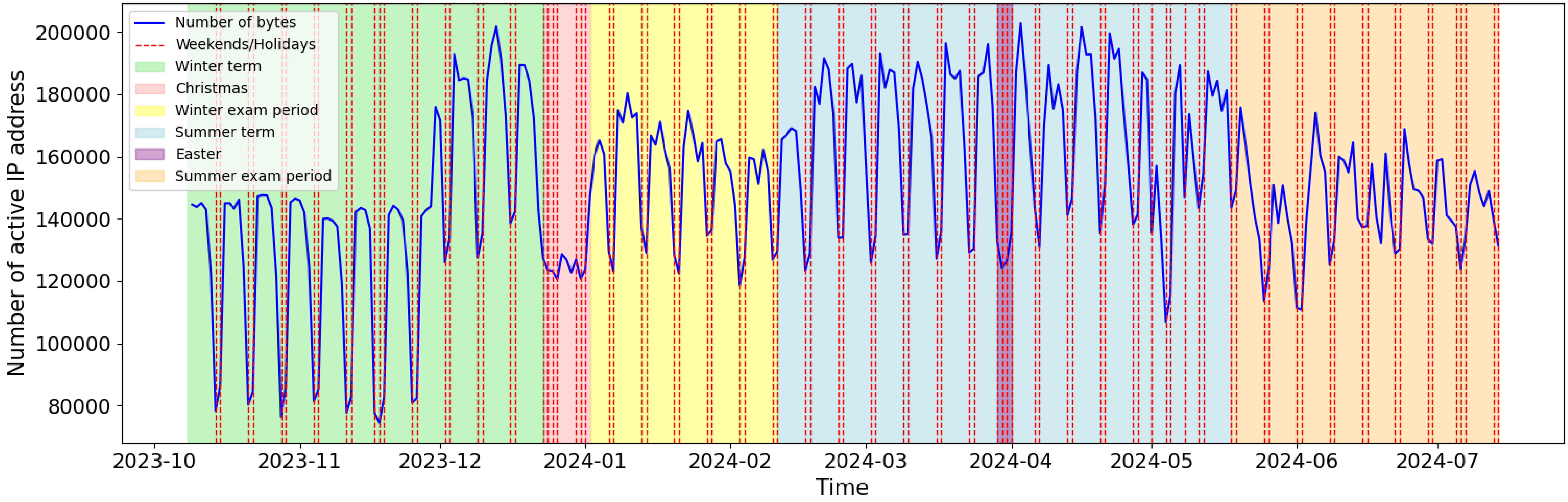
# OBSERVED DATA



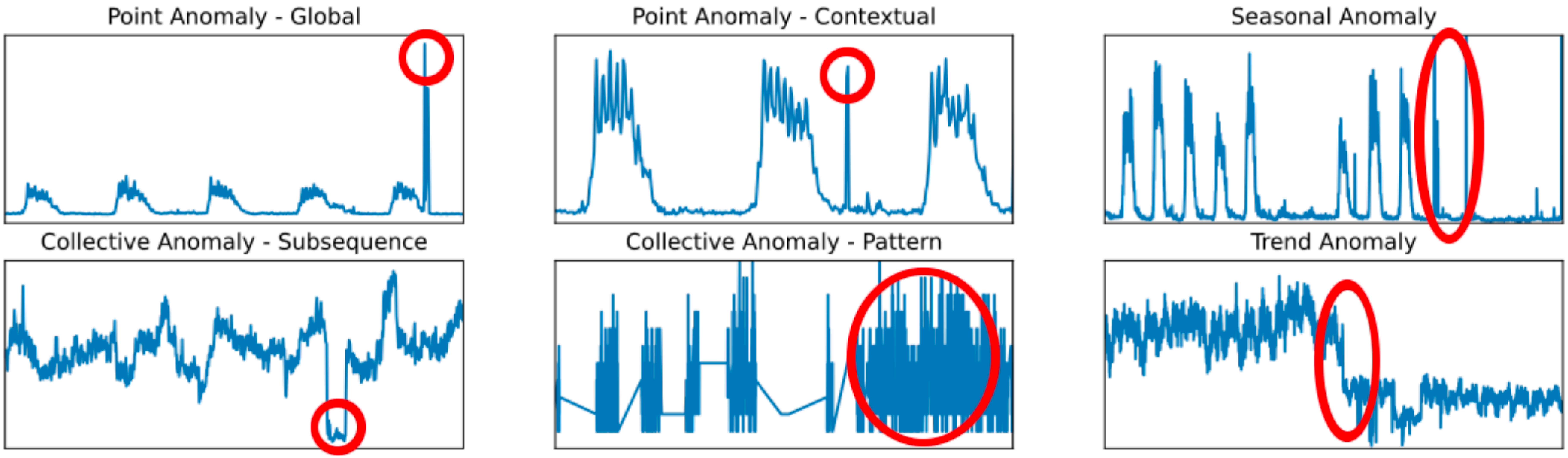
# OBSERVED DATA - INSTITUTION TYPE



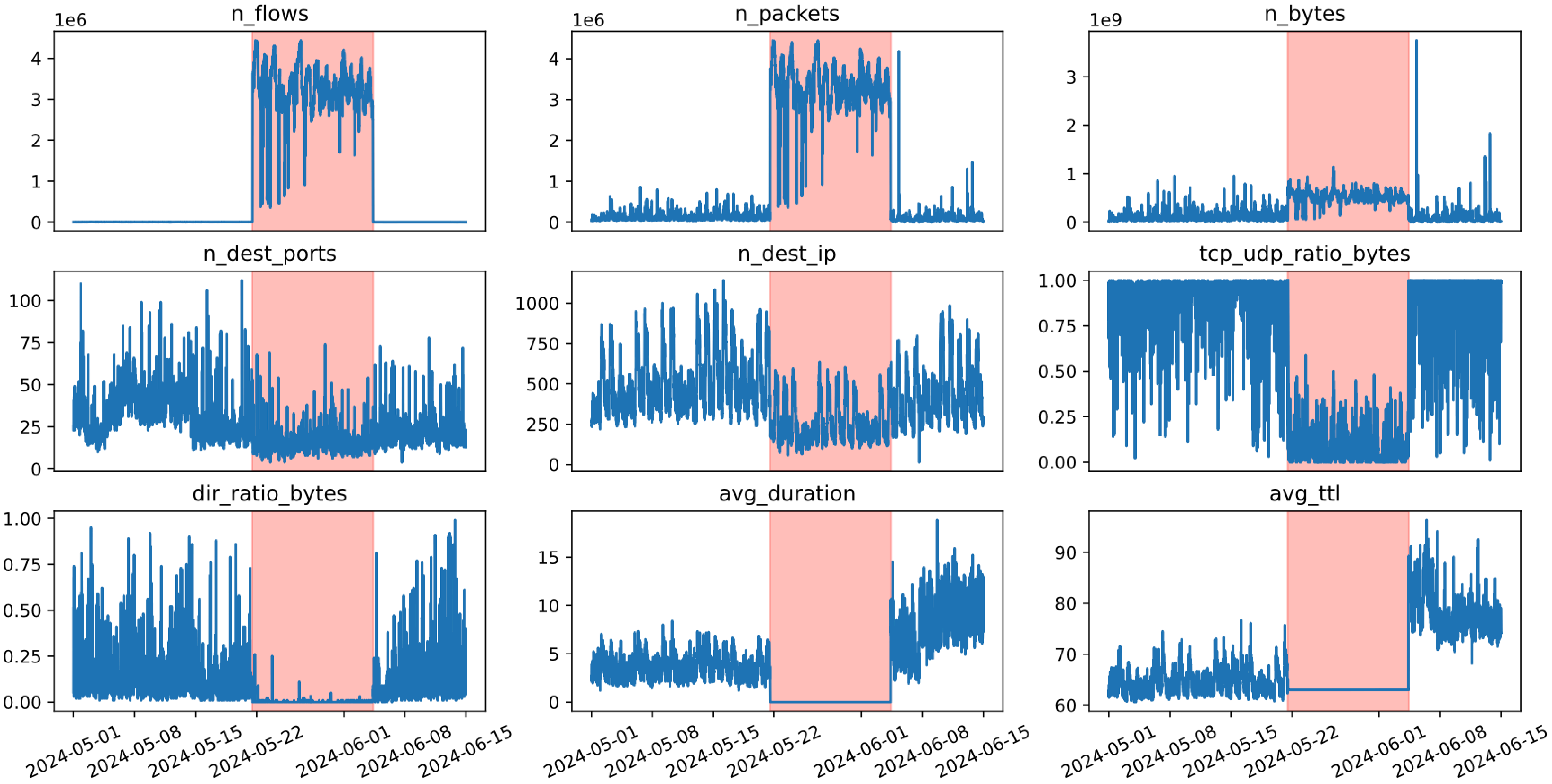
# ACTIVE IP ADDRESSES



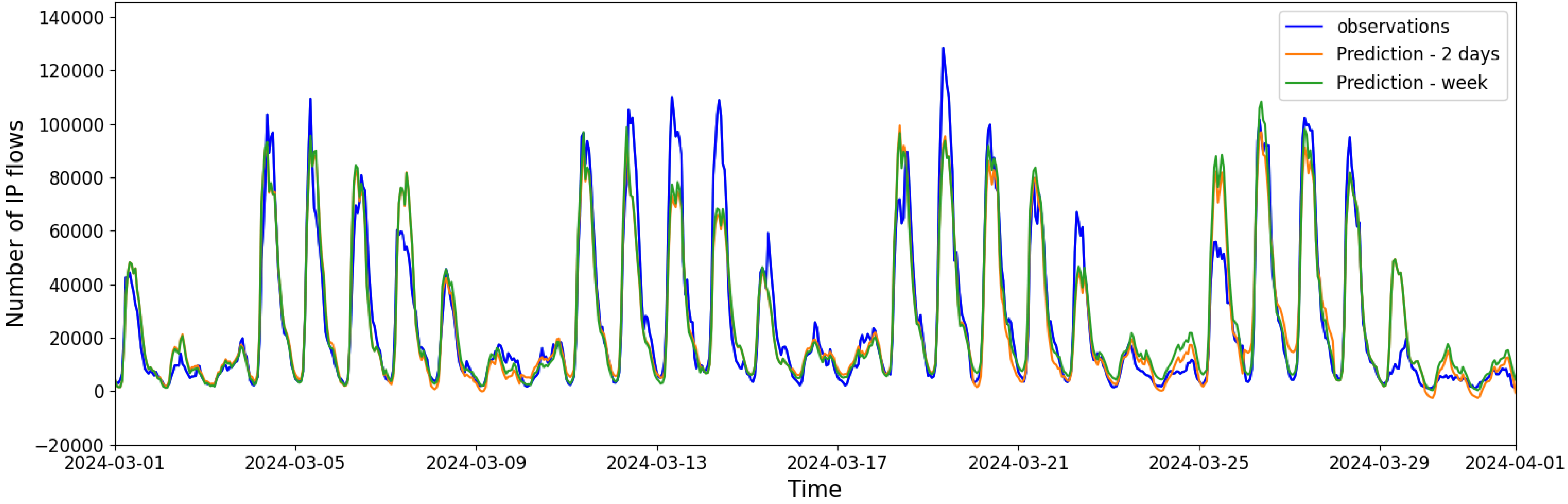
# TYPES OF ANOMALIES



# EXAMPLE: DOS IN DATASET



# MODEL APPLICATION

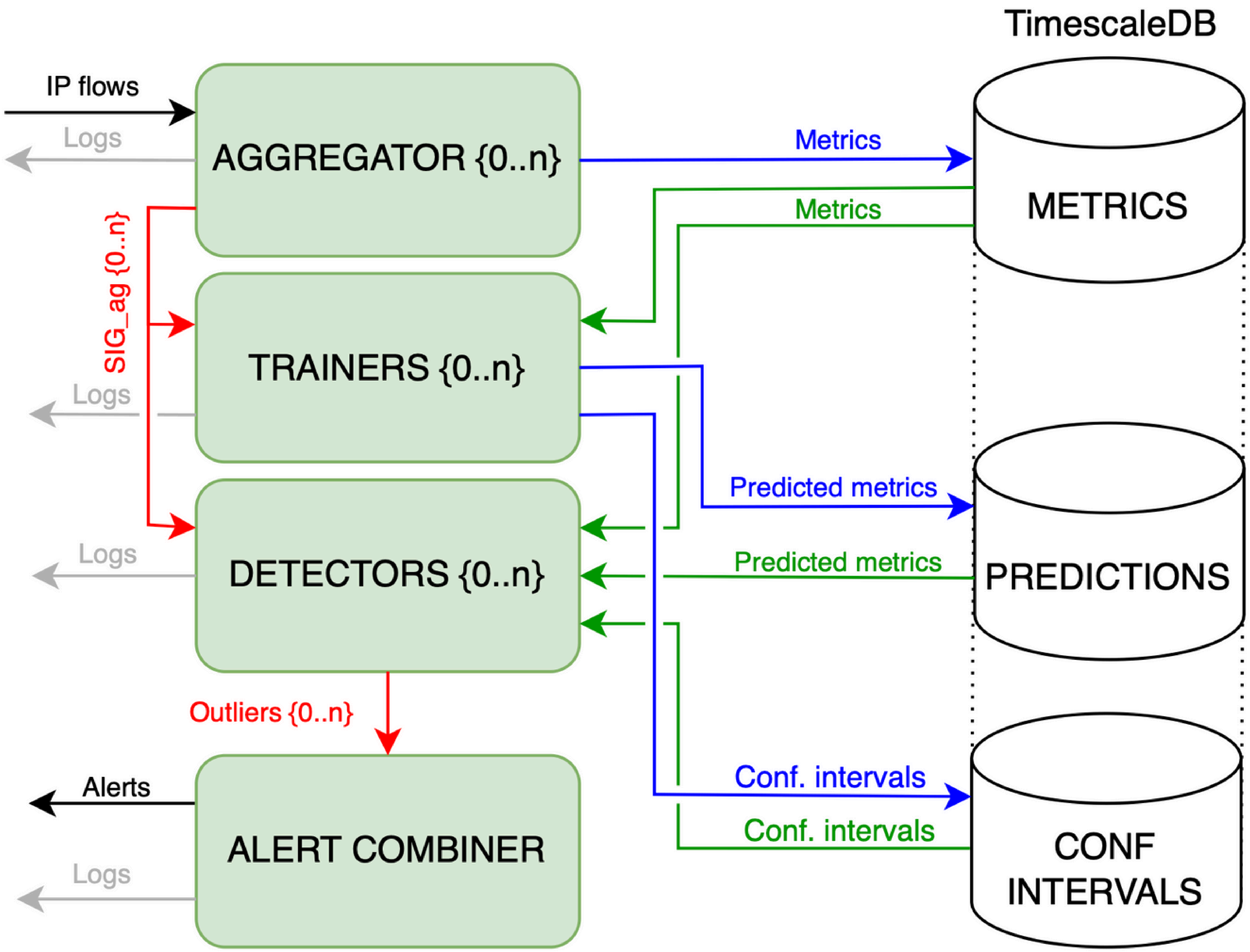


# ANOMALY DETECTION SYSTEM

We propose a novel modular **Network Outlier Detection System (NODS):**

- **built from open-source software**
- enable deployment of anomaly/outlier detection methods based on the forecasting of network traffic in real-world scenarios
- **successfully deployed** on the **real-world ISP network CESNET3**

# ANOMALY DETECTION SYSTEM



# OPEN CHALLANGES

- Many forecasting models can be computationally intensive, making it difficult to scale them for large datasets or in real-time applications. **Optimizing models for efficiency while maintaining accuracy is crucial**.
- Alerts should not only indicate that an anomaly has occurred but also provide context. **Understanding the potential causes and implications of an anomaly is essential** for effective response and mitigation.
- In real-world systems, multiple anomalies can occur simultaneously. Developing methods to assess the **correlation between different alerts and understand their combined impact is crucial for prioritizing responses**.

Dataset

Thank you!

Web page