perfSONAR data analysis of PMP

Ljubomir Hrboka

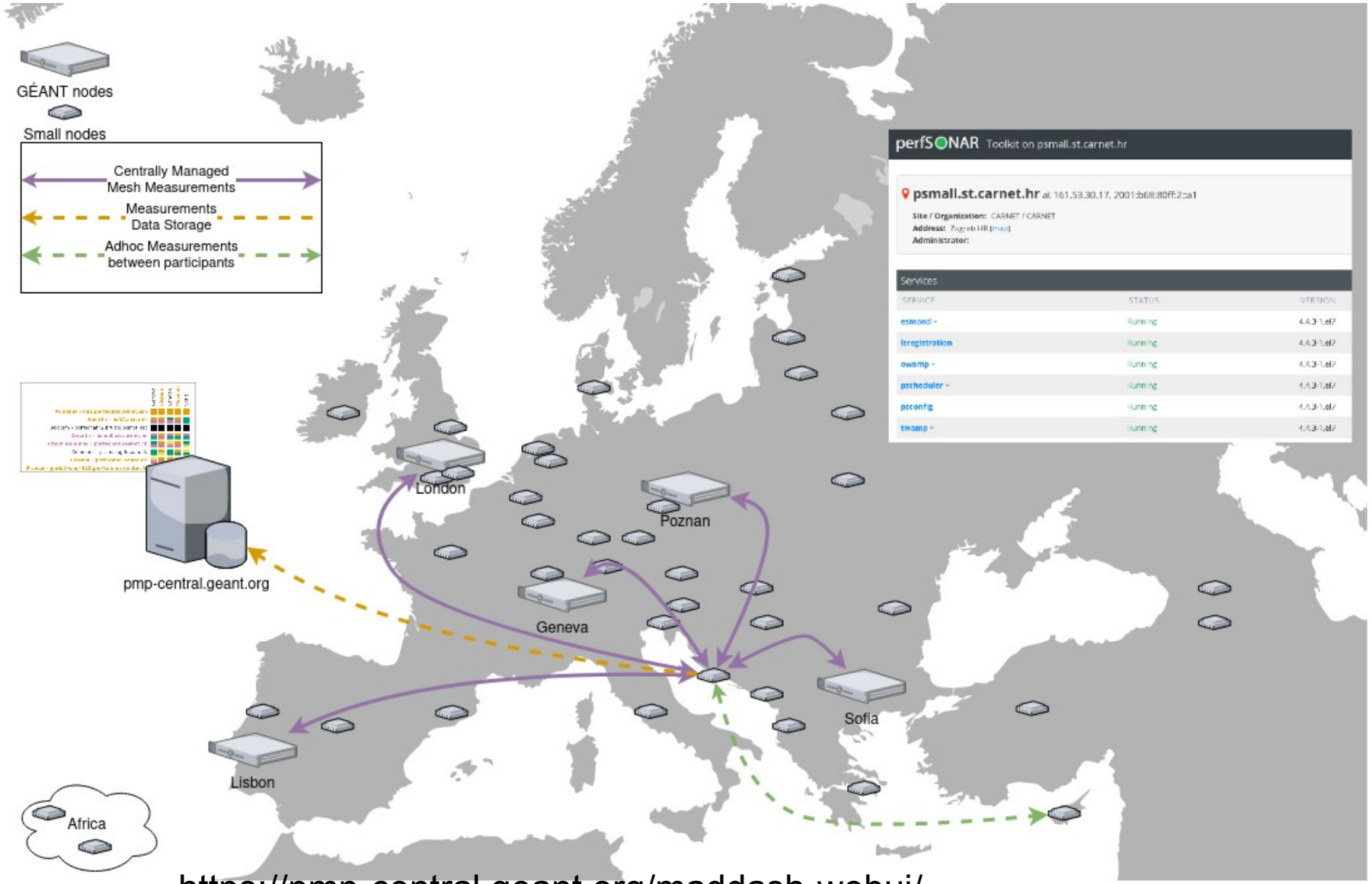
GN4-3 Work Package 6 Task 3 - Monitoring and Management

3rd European perfSONAR User Workshop
2022-05-24

Public

www.geant.org

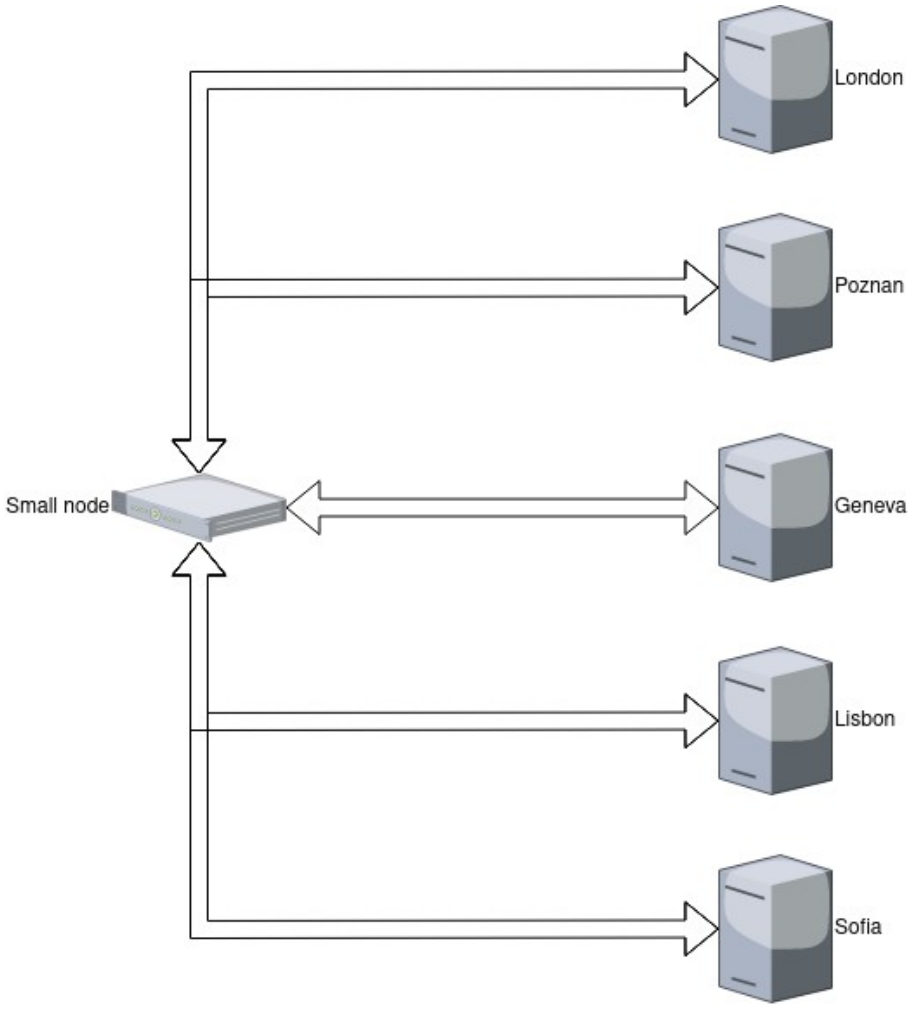
GÉANT Performance Management Platform - PMP



https://pmp-central.geant.org/maddash-webui/

Performance Measurement Platform explores performance to/from the GÉANT backbone while experiencing perfSONAR on small nodes

PMP node-server 2 way measurements



Latency - histogram of 600 values every minute

Jitter - 1 value every minute

RTT - 5 values every ten minutes

Throughput - four values a day

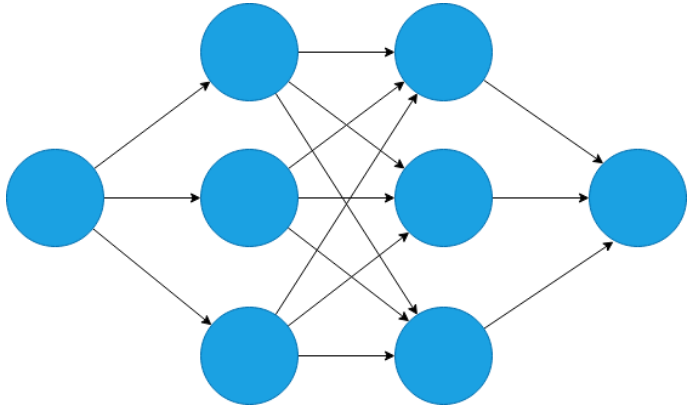
HTTP, DNS, …

Central archive at pmp-central.geant.org

- Central measurement data storage allows usage of ML algorithms to achieve:
  - Holistic view of network performance
  - Detection of barely perceptible or imperceptible anomalies like slight degradation in latency or jitter
  - Detection of deteriorating conditions on multiple links occurring without alarm being triggered
  - Improvement of root cause analysis

The Goal

- Develop a ML model that would be able to detect network anomalies in order to:
  - Pinpoint network areas with ongoing issues
  - Facilitate network planning
  - Support sensitive and/or high data traffic

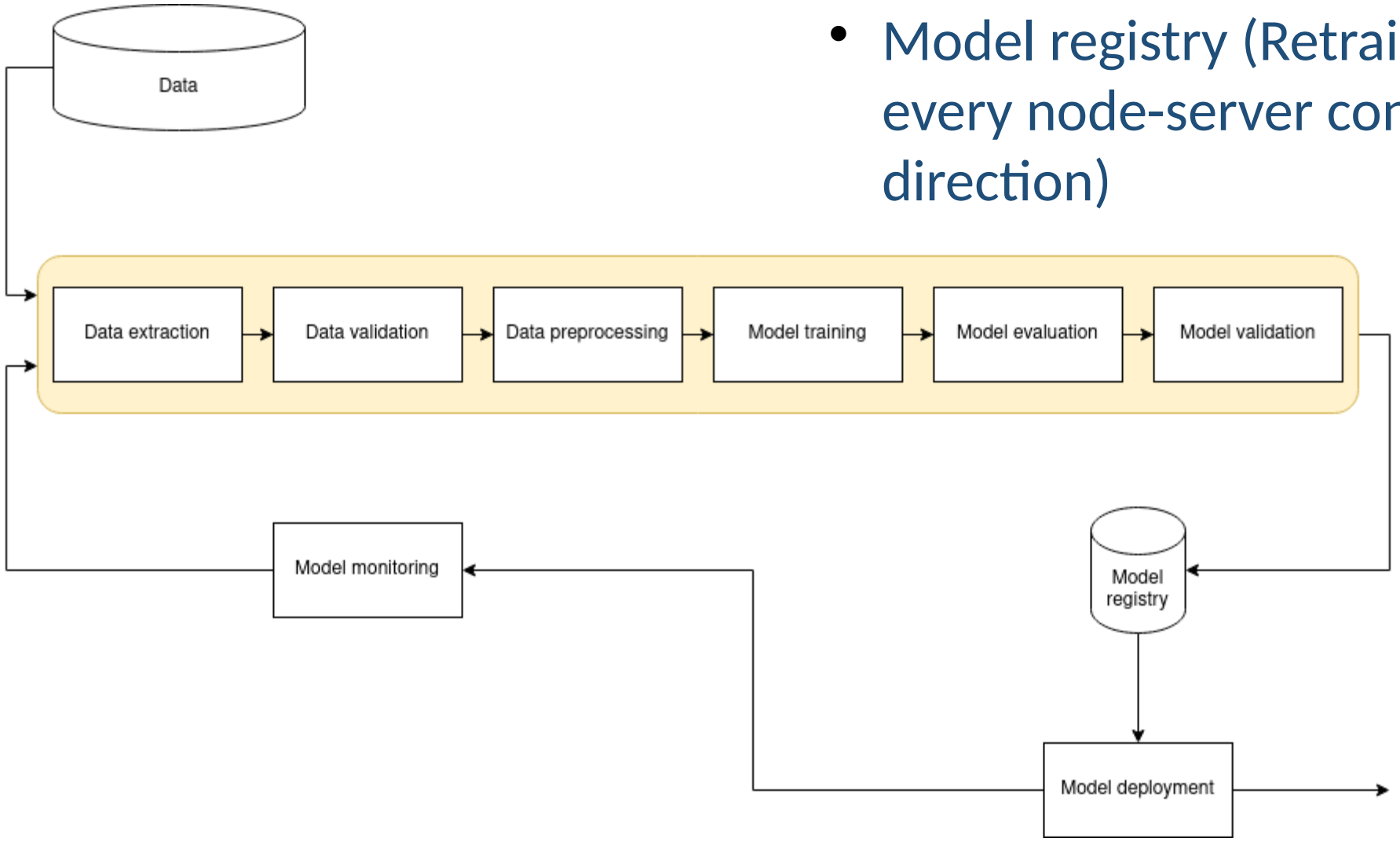


Current Work On Machine Learning Model Development

- Data collection
- Data preprocessing
- Choosing a model
- Model Training
- Model Evaluation
- Model Deployment

- Data Analysis is being performed on the real performance measurement data
- Data Visualization is used for easier understanding and interpretation of relations between the data

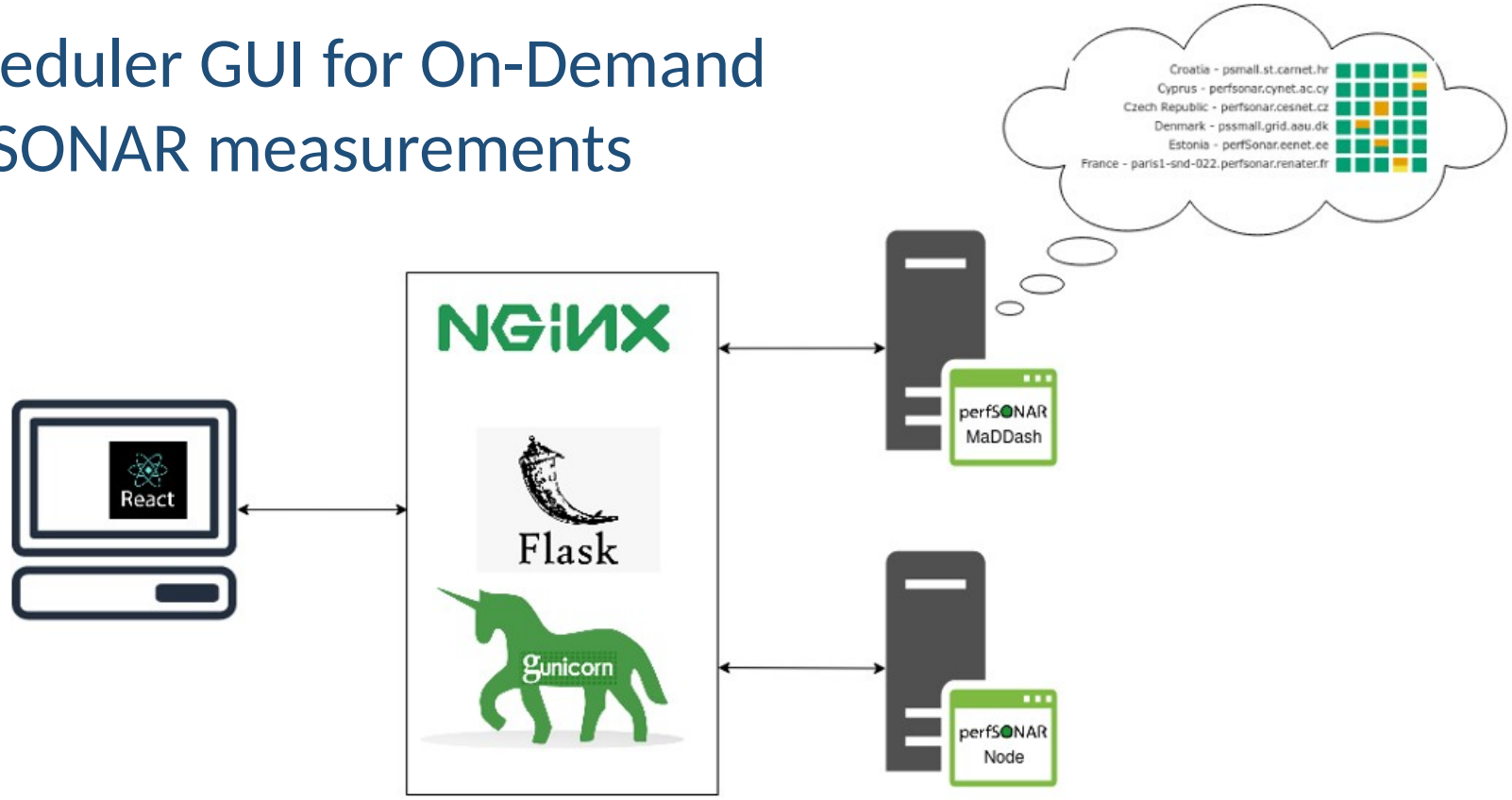
End Goal - ML Automated Pipeline



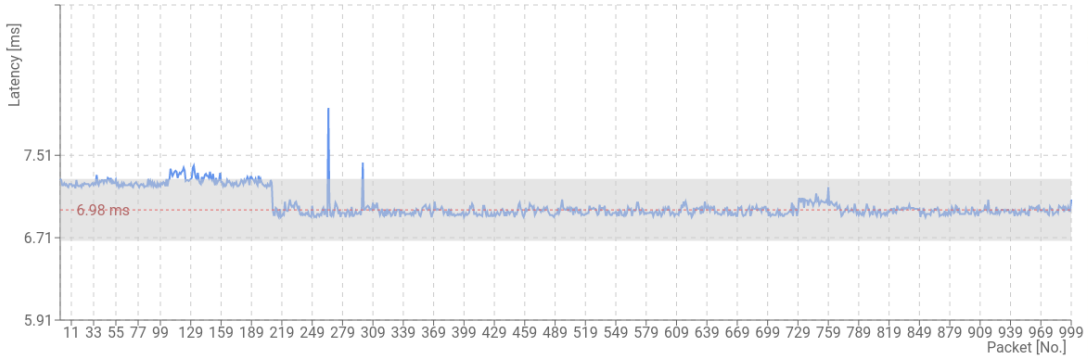
- Model registry (Retrained models for every node-server connection for every direction)

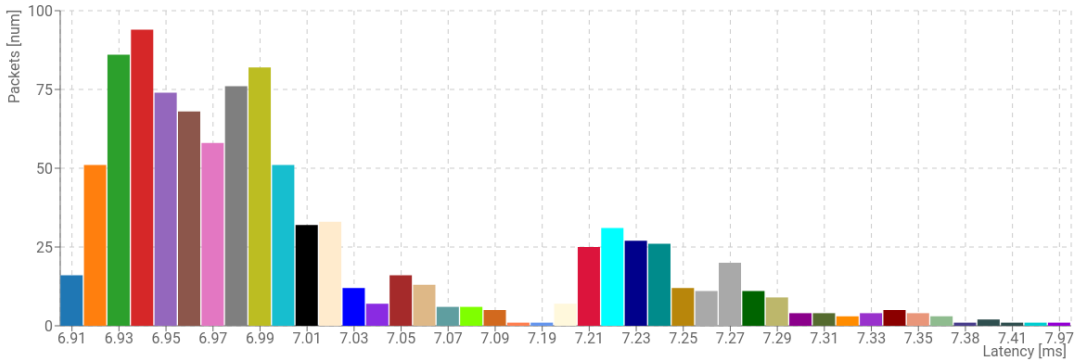
Testing environment for model deployment (psGUI)

Pscheduler GUI for On-Demand perfSONAR measurements

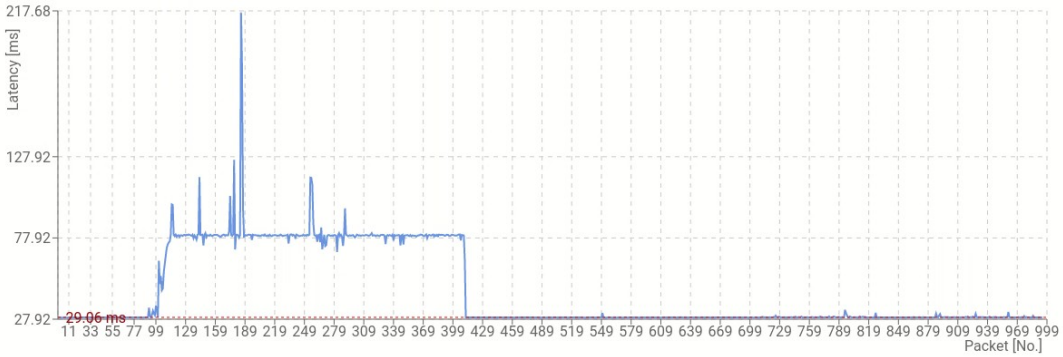


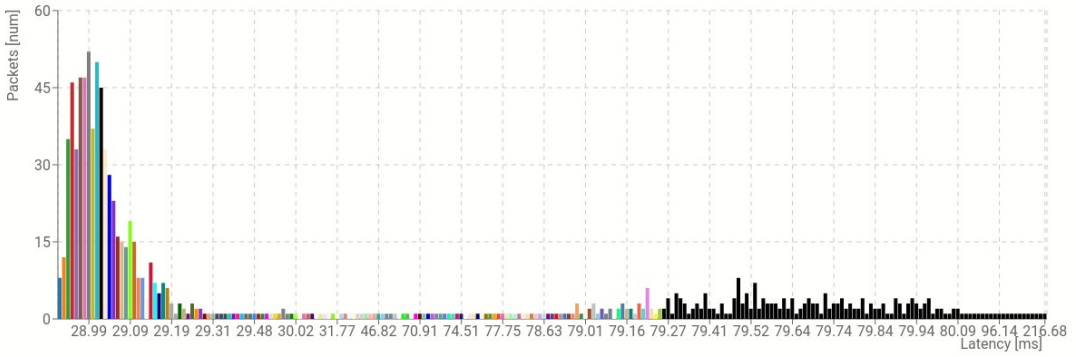
Histogram-owdelay example





Normal

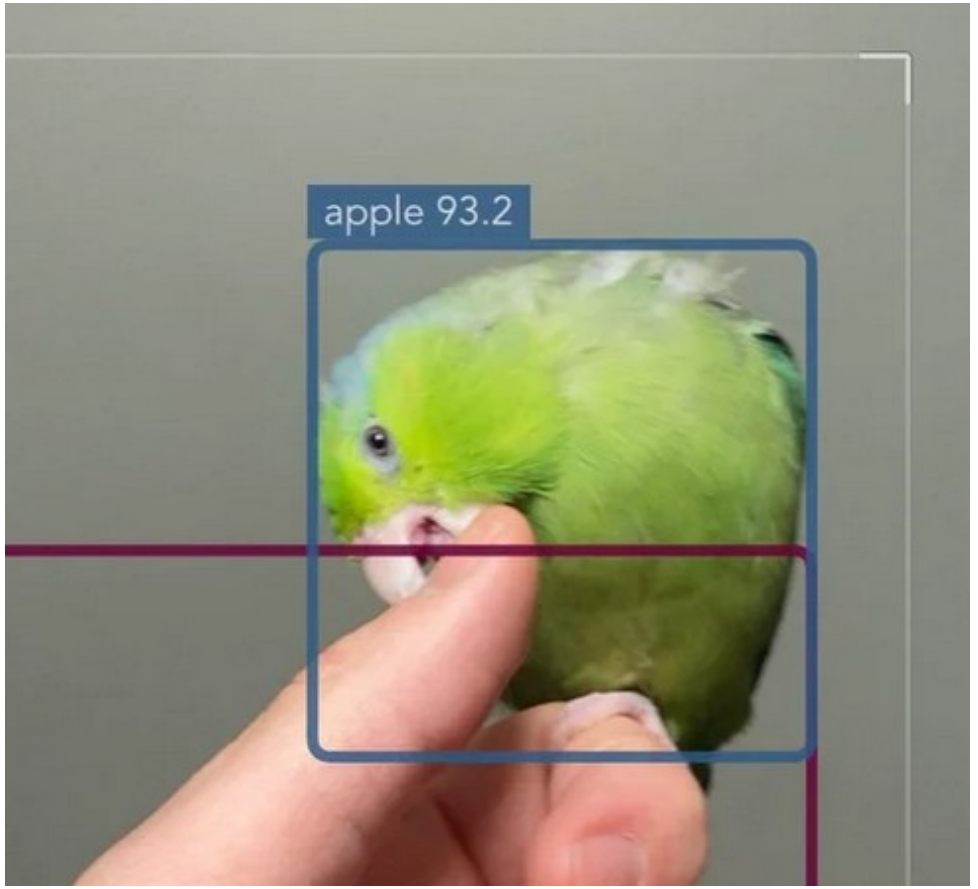




Outlier

The Importance Of Data Analysis

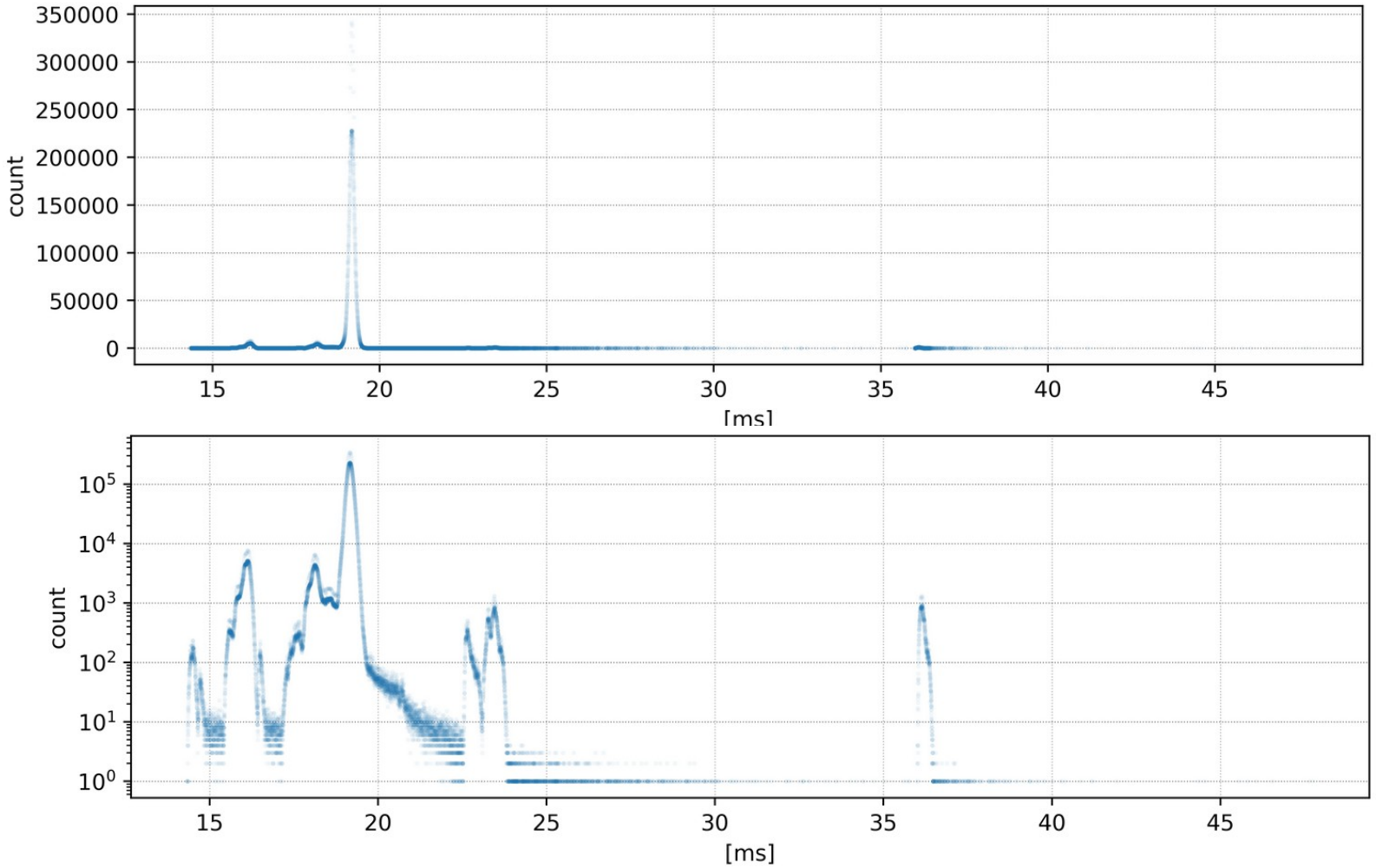
- Garbage In, Garbage Out principle
- Exploratory Data Analysis
  - Observing
  - Categorizing
  - Find missing values
  - Find outliers
  - Correlate
  - Visualize



Procedures Overview

- Data format?
  - JSON files
  - Creating data pipeline
  - Python oriented architecture (NumPy library)
- Labeled Data?
  - No labeled data available
  - Unsupervised learning

Latency Distribution



Period of 2 months;

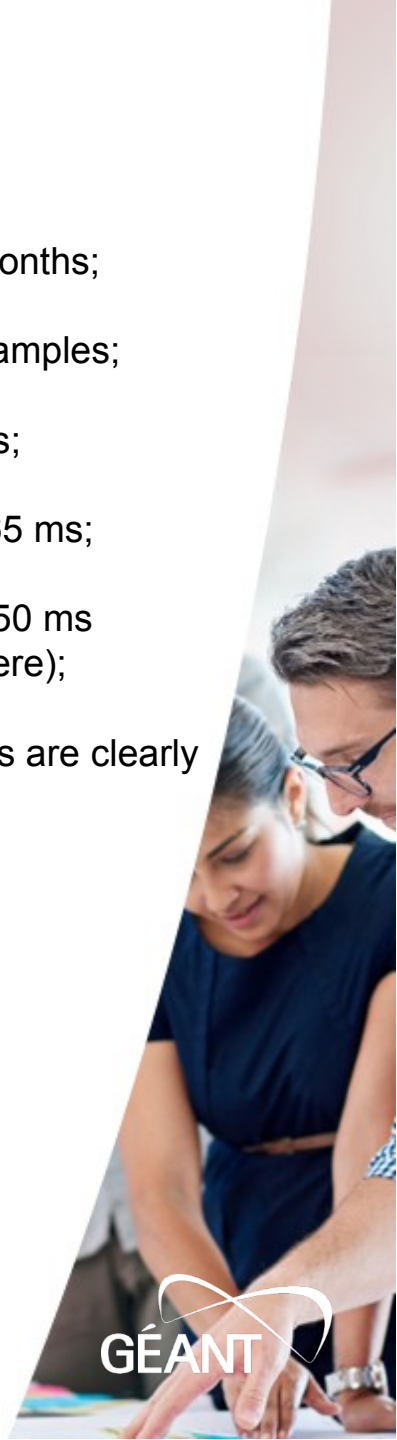
13 011 970 samples;

Min: 14.34 ms;

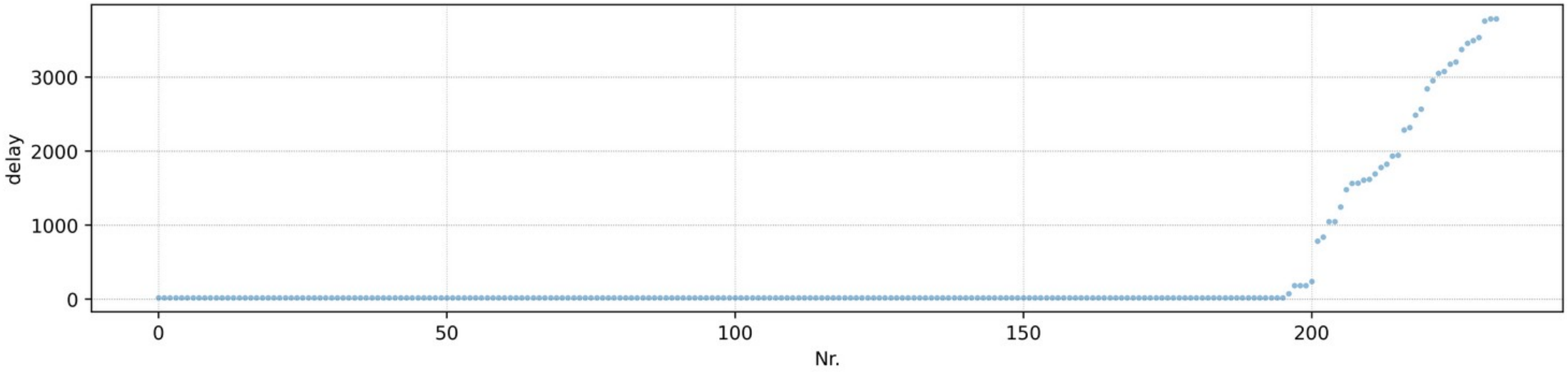
Max: 3789.165 ms;

51 sample > 50 ms
(not shown here);

Distinct spikes are clearly visible.



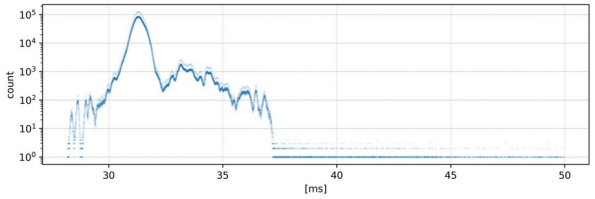
Measurement Error or Anomaly

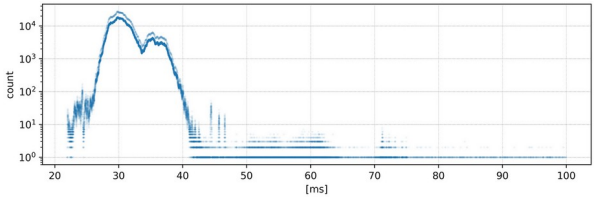


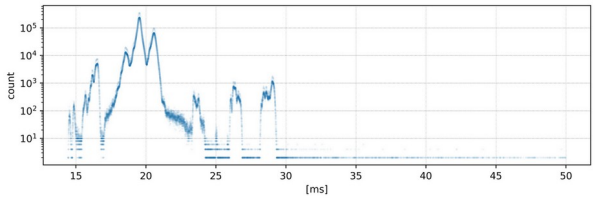
Timestamp: 1627590284 (29.07.2021. 20:24:44)

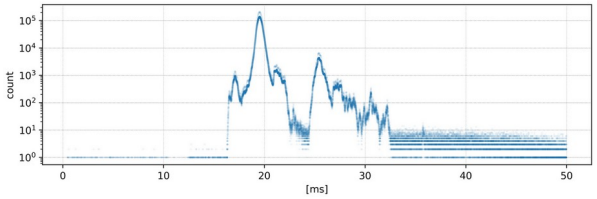
Histogram from 600 packets with 466 unique values within 1 minute.

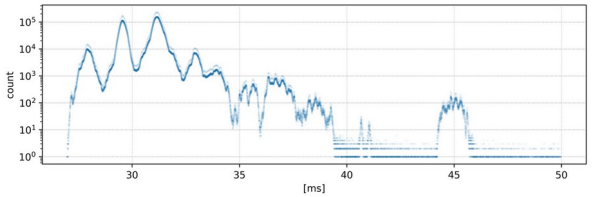
Different Links

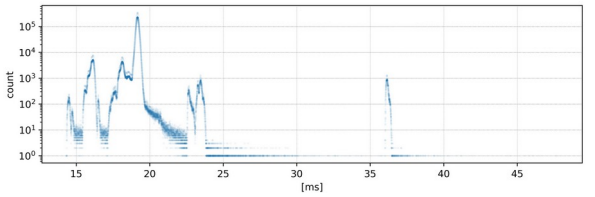




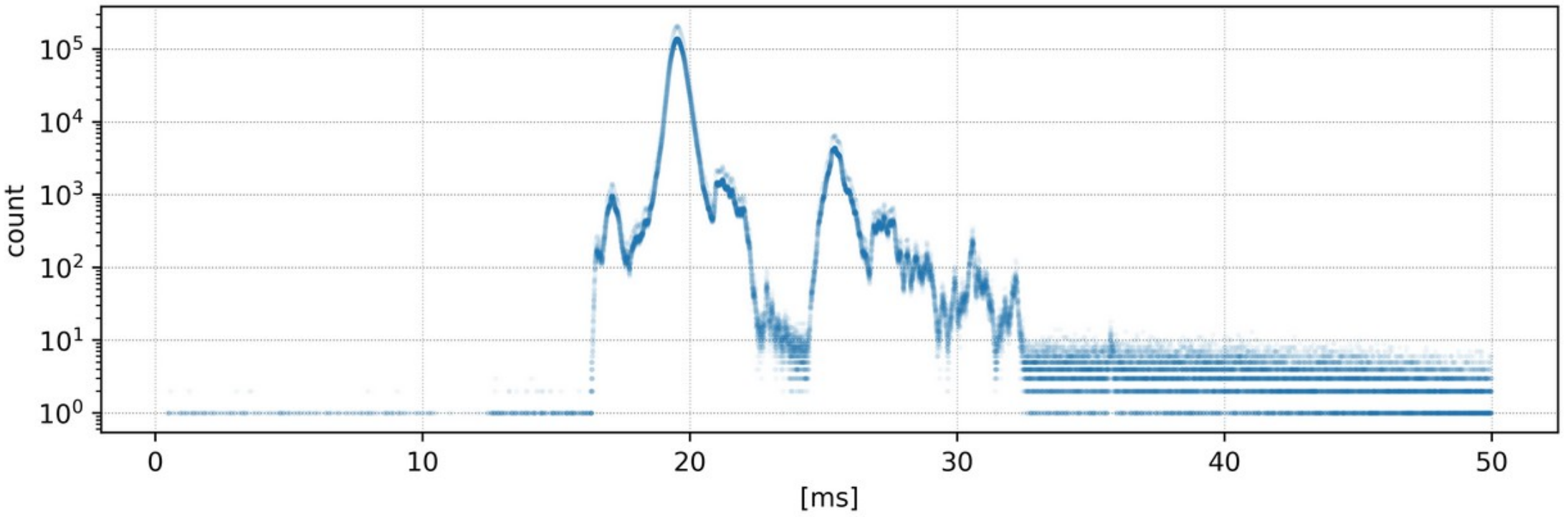




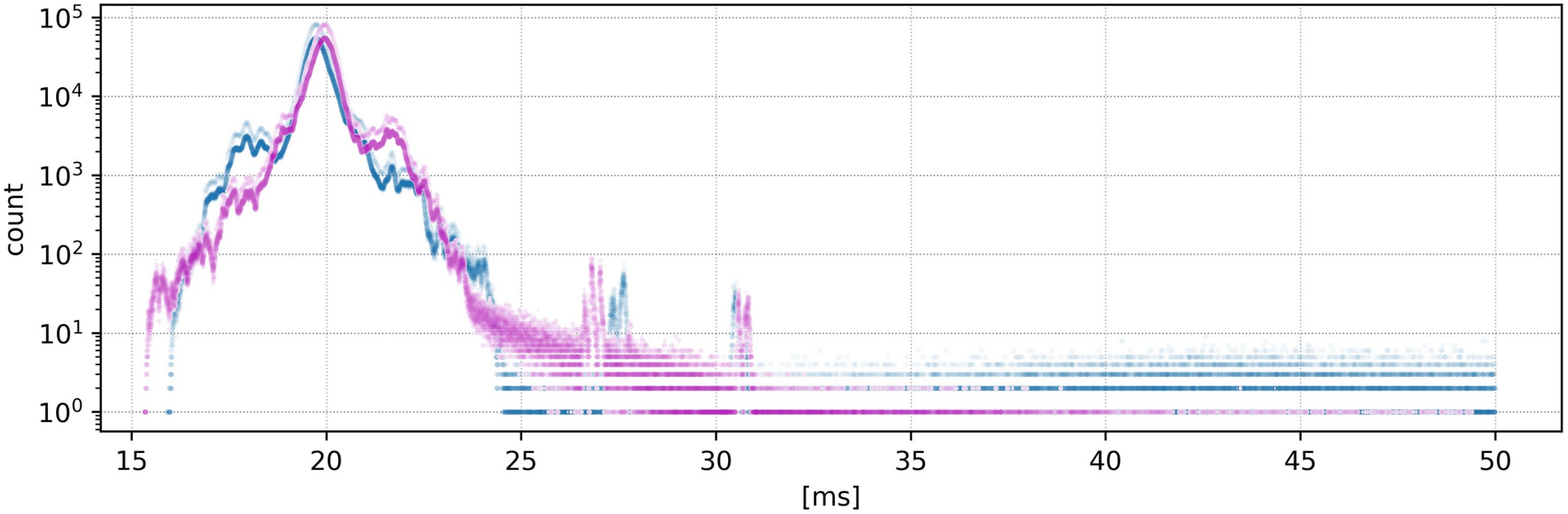




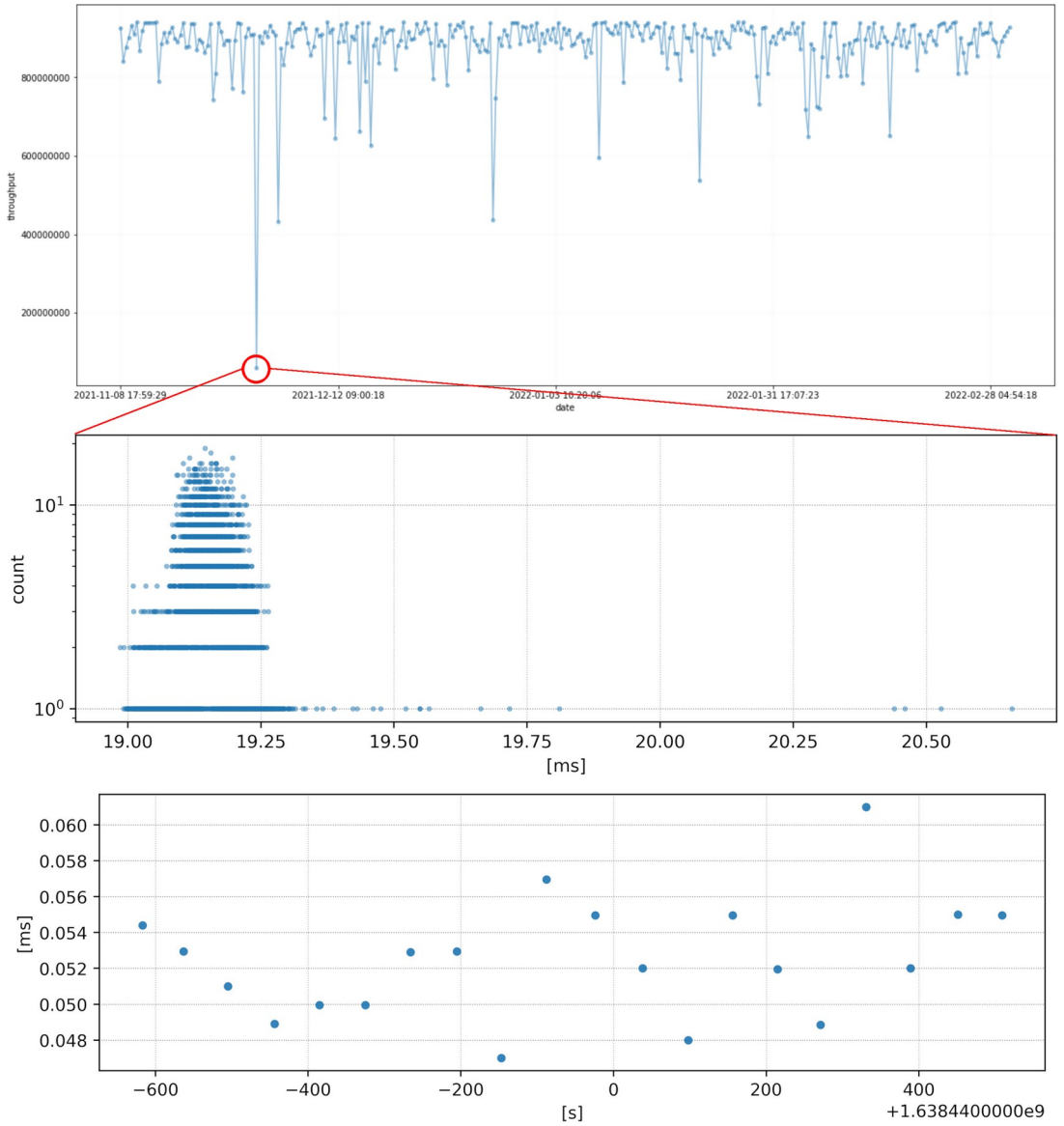
Measurement Errors



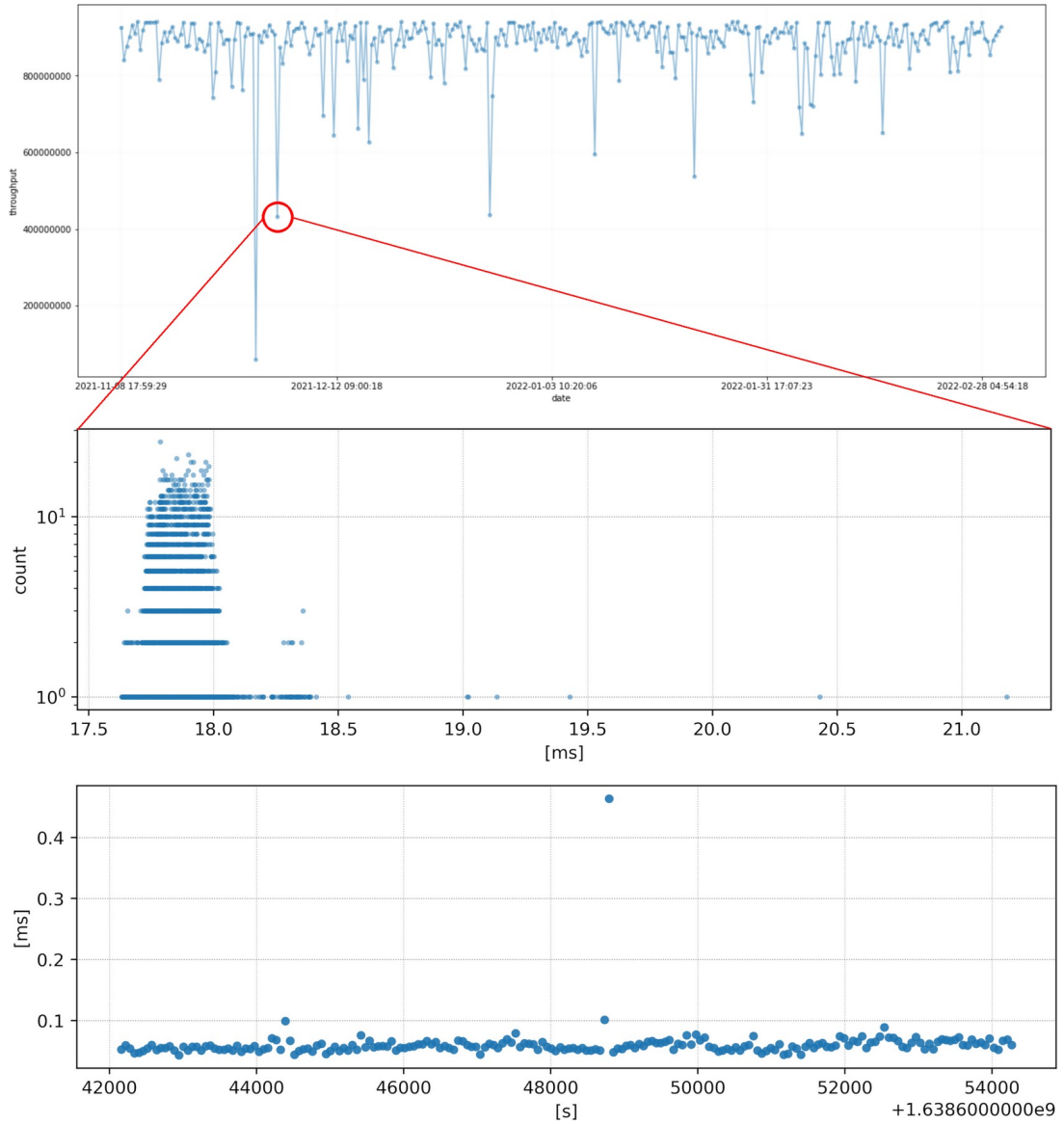
OWD: Same Endpoints - Different Directions



Throughput – Delay – Jitter Outlier (No Correlation)



Throughput – Delay – Jitter Outlier (Correlation)



Current and Future Work

- Test and compare results of several known change detection algorithms on collected data
- Test and deploy ML model for online change detection on collected data
- Create system capable of correlating results from multiple measurements

GÉANT
Networks • Services • People

Thank you

Any questions?

ljubomir@carnet.hr

www.geant.org

The research leading to these results has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No. 856726 (GN4-3).