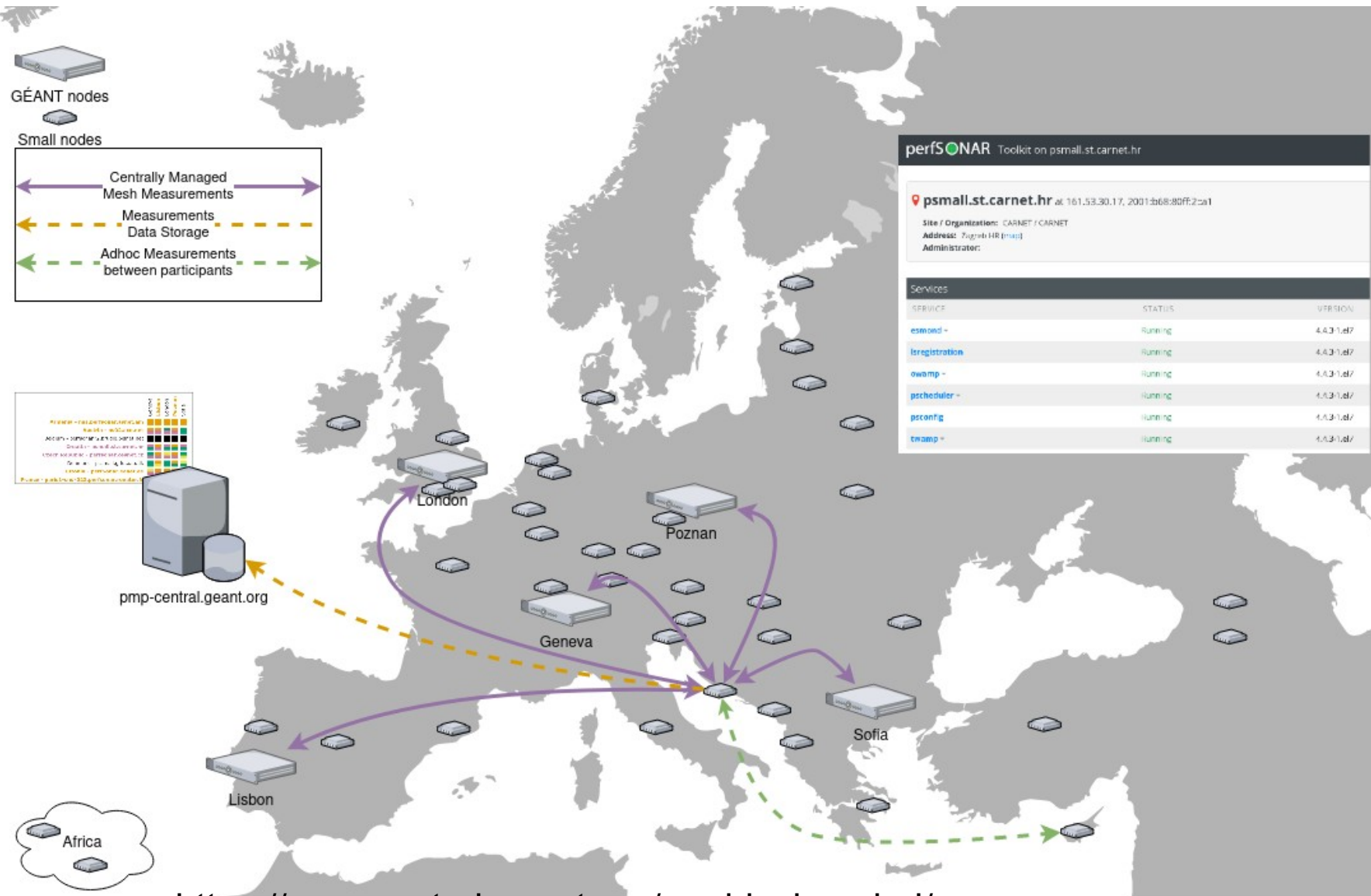# PMP Data Analytics

Ljubomir Hrboka

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

2nd Performance Management Workshop
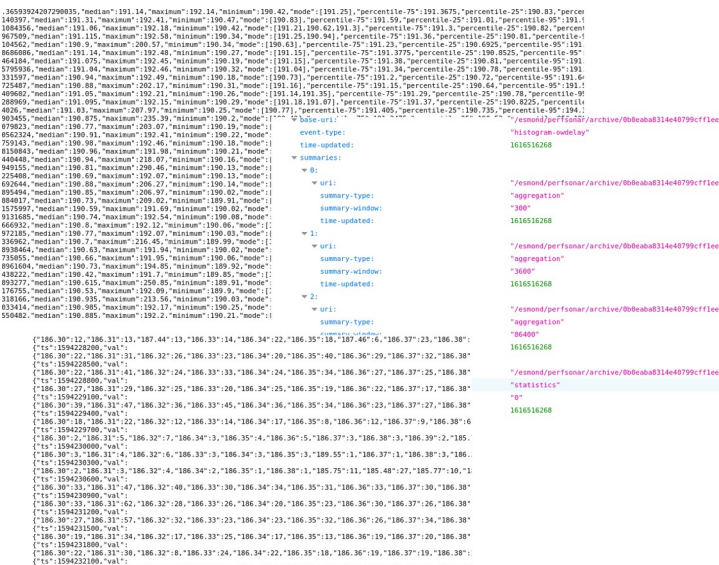2022-03-08

Public

www.geant.org

GÉANT Performance Management Platform - PMP

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

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

www.geant.org

PMP Data Sets (pmp-central.geant.org)

- Latency

- Jitter

- RTT

- Throughput

- HTTP

- DNS

- Couple of years of measurements available

- Data available in JSON format using esmond REST API

.3659392420729003S,"median":191.14,"maximum":192.14,"minimum":190.42,"mode":[191.25],"percentile-75":191.3675,"percentile-25":190.83,"perce
140397,"median":191.31,"maximum":192.41,"minimum":190.47,"mode":[190.83],"percentile-75":191.59,"percentile-25":191.01,"percentile-95":191.9
1084356,"median":191.06,"maximum":192.18,"minimum":190.42,"mode":[191.21,190.62,191.3],"percentile-75":191.3,"percentile-25":190.82,"percent
967509,"median":191.115,"maximum":192.58,"minimum":190.34,"mode":[191.25,190.94],"percentile-75":191.36,"percentile-25":190.81,"percentile-
104562,"median":190.9,"maximum":200.57,"minimum":190.34,"mode":[190.63],"percentile-75":191.3775,"percentile-25":190.8525,"percentile-95":191
8686086,"median":191.14,"maximum":192.48,"minimum":190.27,"mode":[191.15],"percentile-75":191.3775,"percentile-25":190.8525,"percentile-95"
464184,"median":191.075,"maximum":192.45,"minimum":190.19,"mode":[191.15],"percentile-75":191.38,"percentile-25":190.81,"percentile-95":191
5795936,"median":191.04,"maximum":192.46,"minimum":190.32,"mode":[191.04],"percentile-75":191.34,"percentile-25":190.78,"percentile-95":191
331597,"median":190.94,"maximum":192.49,"minimum":190.18,"mode":[190.73],"percentile-75":191.2,"percentile-25":190.72,"percentile-95":191.6
725487,"median":190.88,"maximum":202.17,"minimum":190.31,"mode":[191.16],"percentile-75":191.15,"percentile-25":190.64,"percentile-95":191.6
409602,"median":191.05,"maximum":192.21,"minimum":190.26,"mode":[191.14,191.35],"percentile-75":191.29,"percentile-25":190.78,"percentile-9
288969,"median":191.095,"maximum":192.15,"minimum":190.29,"mode":[191.18,191.07],"percentile-75":191.37,"percentile-25":190.8225,"percentile
4026,"median":191.03,"maximum":207.97,"minimum":190.25,"mode":[190.77],"percentile-75":191.405,"percentile-25":190.735,"percentile-95":194.1
903455,"median":190.875,"maximum":235.39,"minimum":190.2,"mode":[
879823,"median":190.77,"maximum":203.07,"minimum":190.19,"mode":[
8562324,"median":190.91,"maximum":192.41,"minimum":190.22,"mode":[
759143,"median":190.98,"maximum":192.46,"minimum":190.18,"mode":[
8150843,"median":190.96,"maximum":191.98,"minimum":190.21,"mode":[
440448,"median":190.94,"maximum":218.07,"minimum":190.16,"mode":[
949155,"median":190.81,"maximum":290.46,"minimum":190.13,"mode":[
225408,"median":190.69,"maximum":192.07,"minimum":190.13,"mode":[
692644,"median":190.88,"maximum":206.27,"minimum":190.14,"mode":[
895494,"median":190.85,"maximum":206.97,"minimum":190.02,"mode":[
848017,"median":190.73,"maximum":209.02,"minimum":189.91,"mode":[
1575997,"median":190.59,"maximum":191.69,"minimum":190.02,"mode":[
9131685,"median":190.74,"maximum":192.54,"minimum":190.08,"mode":[
666932,"median":190.8,"maximum":192.12,"minimum":190.06,"mode":[
972185,"median":190.77,"maximum":192.07,"minimum":190.03,"mode":[
336962,"median":190.7,"maximum":216.45,"minimum":189.99,"mode":[
8938464,"median":190.63,"maximum":191.94,"minimum":190.02,"mode":[
735055,"median":190.66,"maximum":191.95,"minimum":190.06,"mode":[
8961604,"median":190.73,"maximum":194.85,"minimum":189.92,"mode":[
438222,"median":190.42,"maximum":191.7,"minimum":189.85,"mode":[
893277,"median":190.615,"maximum":250.85,"minimum":189.91,"mode":[
176755,"median":190.53,"maximum":192.09,"minimum":189.9,"mode":[
318166,"median":190.935,"maximum":213.56,"minimum":190.03,"mode":[
033414,"median":190.985,"maximum":192.17,"minimum":190.25,"mode":[
550482,"median":190.885,"maximum":192.2,"minimum":190.21,"mode":[

base-uri: ...
  event-type: "histogram-owdelay"
  time-updated: 1616516268
  ▼ summaries:
    ▼ 0:
      ▼ uri: "/esmond/perfsonar/archive/0b0eaba8314e40799cff1ee0
        summary-type: "aggregation"
        summary-window: "300"
        time-updated: 1616516268
    ▼ 1:
      ▼ uri: "/esmond/perfsonar/archive/0b0eaba8314e40799cff1ee0
        summary-type: "aggregation"
        summary-window: "3600"
        time-updated: 1616516268
    ▼ 2:
      ▼ uri: "/esmond/perfsonar/archive/0b0eaba8314e40799cff1ee0
        summary-type: "aggregation"

{"186.30":12,"186.31":13,"187.44":13,"186.33":14,"186.34":22,"186.35":18,"187.46":6,"186.37":23,"186.38":
{"ts":1594228200,"val":   "86400"
{"186.30":22,"186.31":31,"186.32":26,"186.33":23,"186.34":20,"186.35":40,"186.36":29,"186.37":32,"186.38" 1616516268
{"186.30":22,"186.31":41,"186.32":24,"186.33":33,"186.34":24,"186.35":34,"186.36":27,"186.37":25,"186.38"
{"ts":1594228500,"val":   "/esmond/perfsonar/archive/0b0eaba8314e40799cff1ee0
{"186.30":27,"186.31":29,"186.32":25,"186.33":20,"186.35":19,"186.36":22,"186.37":17,"186.38" "statistics"
{"ts":1594229100,"val":   "0"
{"186.30":39,"186.31":47,"186.32":36,"186.33":45,"186.34":36,"186.35":34,"186.36":23,"186.37":27,"186.38"
{"ts":1594229400,"val":   1616516268
{"186.30":18,"186.31":22,"186.32":12,"186.33":14,"186.34":17,"186.35":8,"186.36":12,"186.37":9,"186.38":6
{"ts":1594229700,"val":
{"186.30":2,"186.31":5,"186.32":7,"186.34":3,"186.35":4,"186.36":5,"186.37":3,"186.38":3,"186.39":2,"185.
{"ts":1594230000,"val":
{"186.30":3,"186.31":4,"186.32":6,"186.33":3,"186.34":3,"186.35":3,"189.55":1,"186.37":1,"186.38":3,"186.
{"ts":1594230300,"val":
{"186.30":2,"186.31":3,"186.32":4,"186.34":2,"186.35":1,"186.38":1,"185.75":1,"185.48":27,"185.77":10,"1
{"ts":1594230600,"val":
{"186.30":33,"186.31":47,"186.32":40,"186.33":30,"186.34":34,"186.35":31,"186.36":33,"186.37":30,"186.38"
{"ts":1594230900,"val":
{"186.30":33,"186.31":62,"186.32":28,"186.33":26,"186.34":20,"186.35":23,"186.36":30,"186.37":26,"186.38"
{"ts":1594231200,"val":
{"186.30":27,"186.31":57,"186.32":32,"186.33":37,"186.34":23,"186.35":22,"186.36":26,"186.37":34,"186.38"
{"ts":1594231500,"val":
{"186.30":19,"186.31":34,"186.32":17,"186.33":25,"186.34":17,"186.35":13,"186.36":19,"186.37":20,"186.38"
{"ts":1594231800,"val":
{"186.30":22,"186.31":38,"186.32":8,"186.33":24,"186.34":22,"186.35":18,"186.36":19,"186.37":19,"186.38":
{"ts":1594232100,"val":

Detecting Network Anomalies (Outliers) from Network Performance Data

- 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
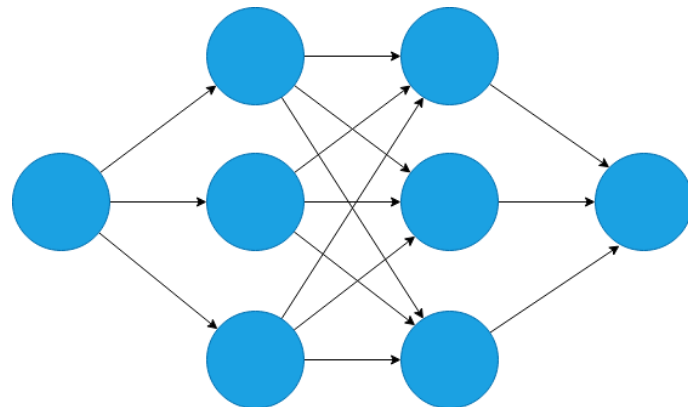
Machine Learning Model Development

- 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

GÉANT

The Importance of Understanding the Data

- Data Pre-processing is the most important part of the Machine Learning algorithm - "Garbage In, Garbage Out" principle

- Exploratory Data Analysis

  - Observing

  - Categorizing

  - Find missing values

  - Find outliers

  - Correlate

  - Visualize

  - ...

apple 93.2

GÉANT
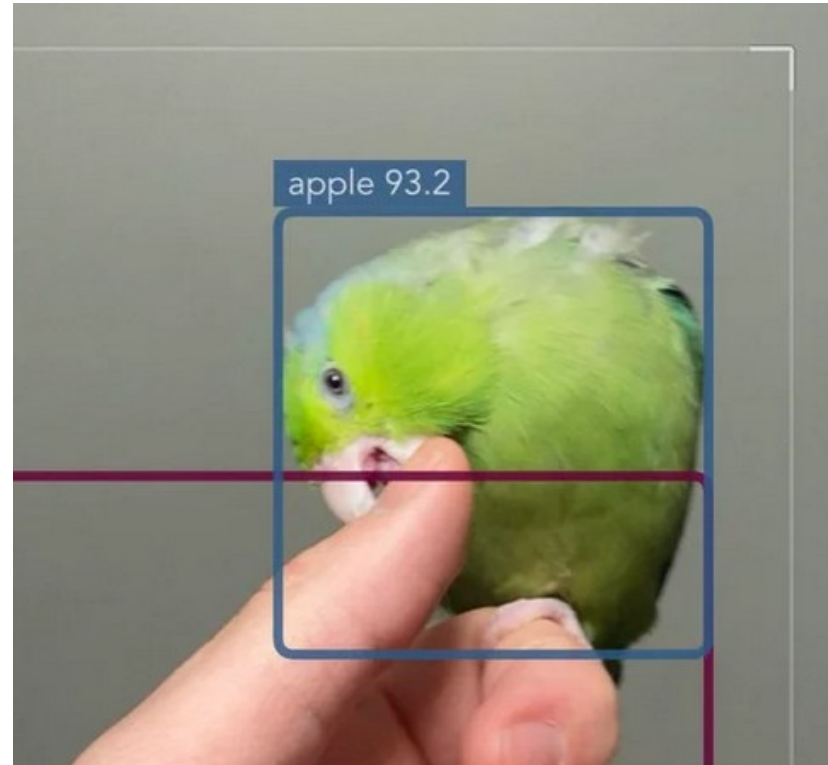
Measurement Data Overview

- What type of data and how much of it is available?

  - One-way Latency (nominally histogram of 600 values every minute) - powstream

  - Jitter (nominally 1 value every minute) - powstream

  - Throughput (nominally four values a day) - iperf3

  - RTT (nominally 5 values every ten minutes) - ping

  - Traceroute (nominally 1 measurement every ten minutes) - traceroute

  - HTTP response (1 measurement every hour) - curl

  - DNS (1 measurement every hour) – dnspy

Procedures Overview

- Data format?

  - JSON files

  - Creating data pipeline

  - Python oriented architecture (NumPy library)

- Labeled Data?

  - No labeled data available

  - Unsupervised learning

Testing environment for model deployment (psGUI)

Pscheduler GUI for On-Demand perfSONAR measurements
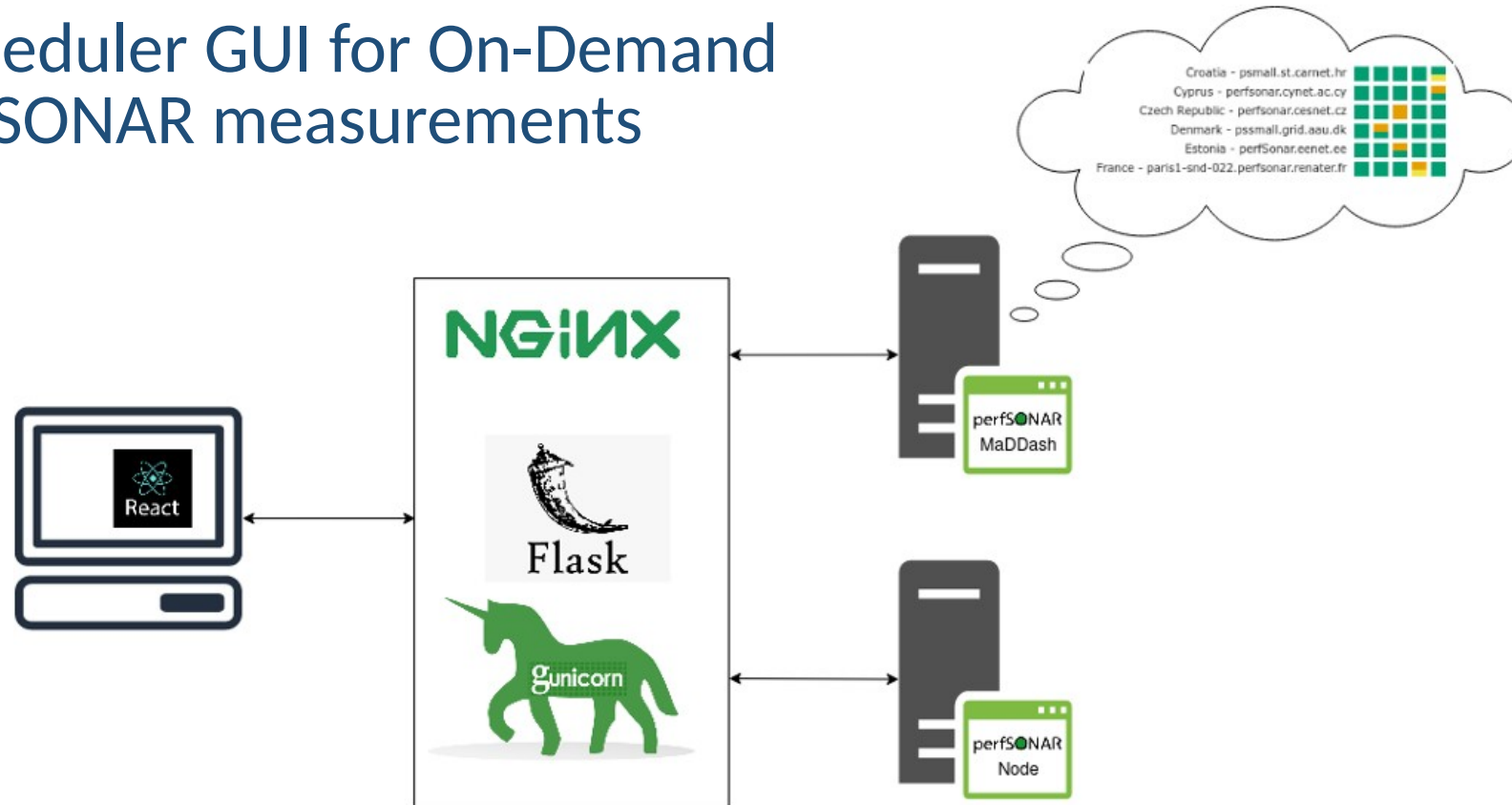
Histogram-owdelay example
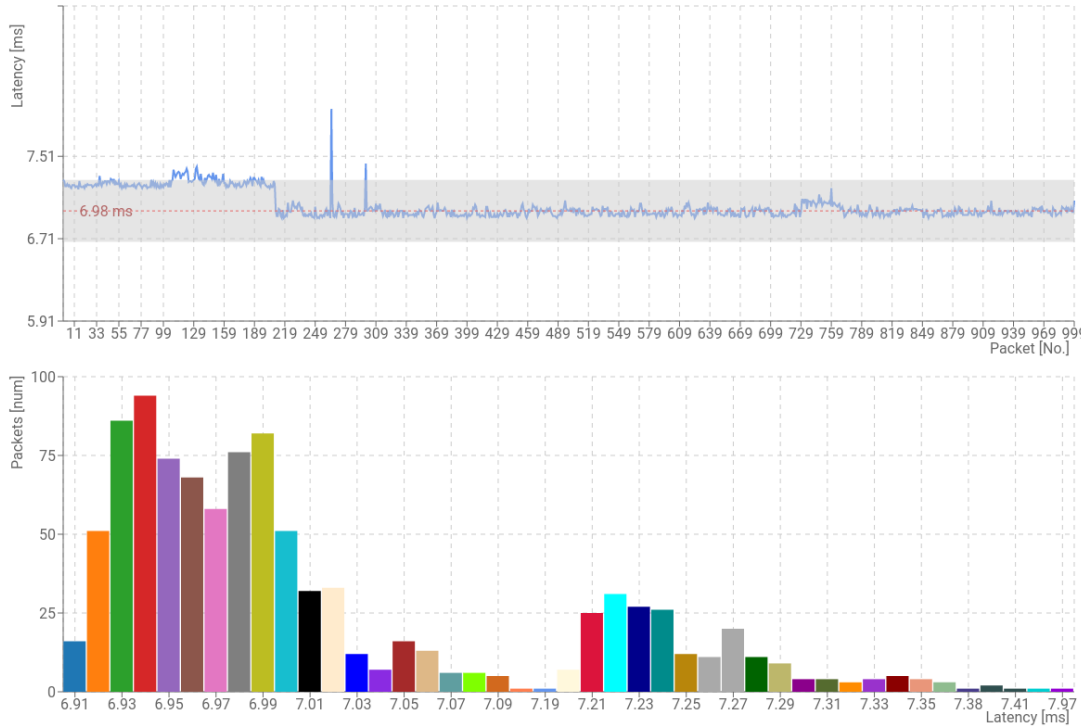
"Normal"
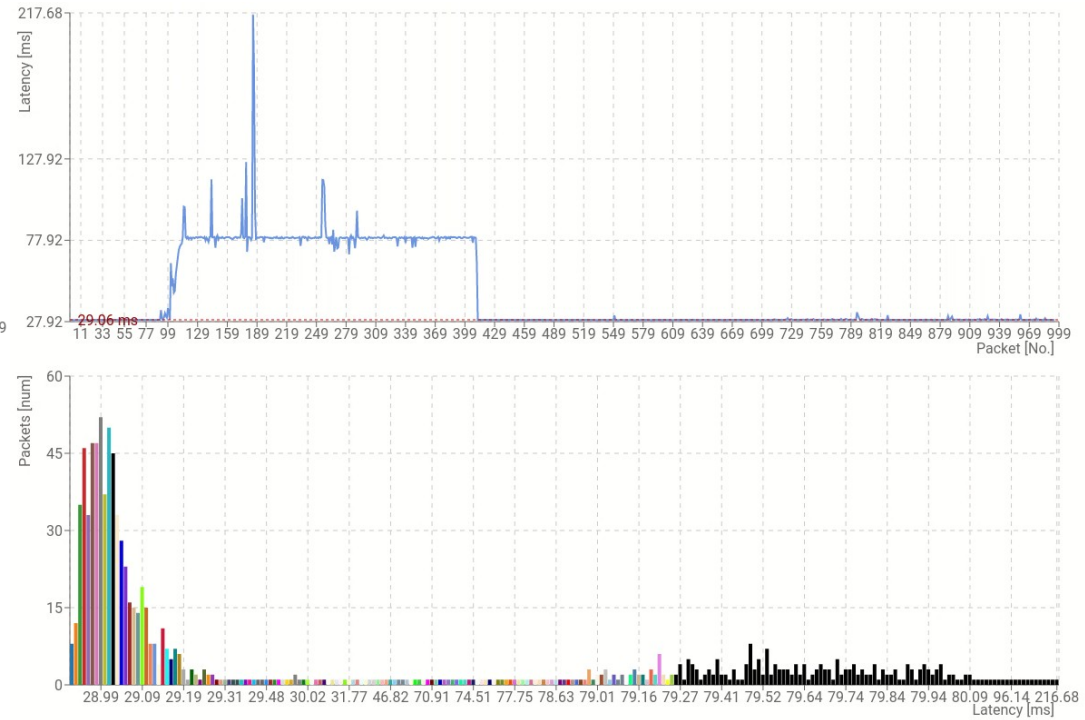
"Anomalous"

Latency Distribution
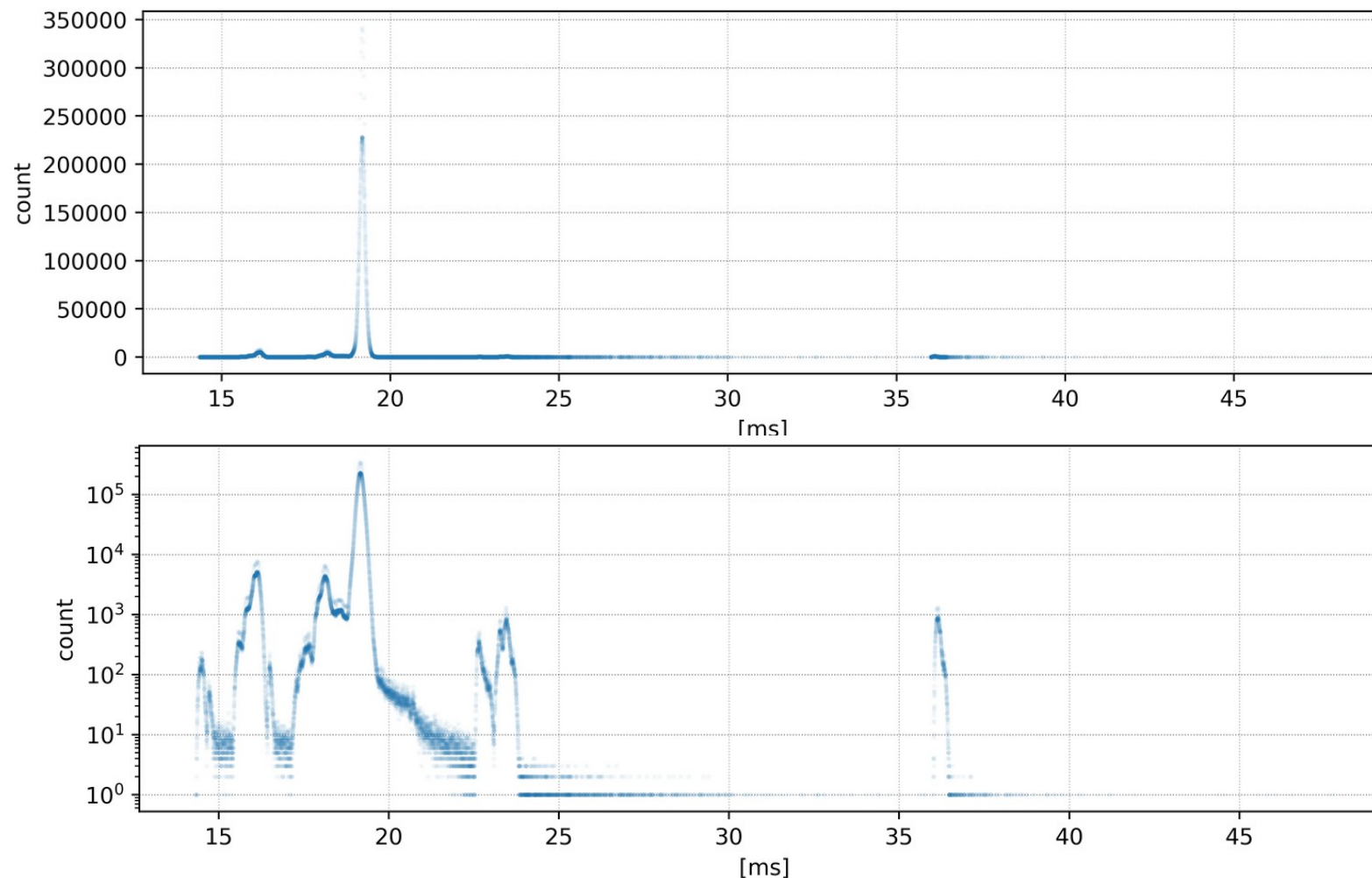
Period of 2 months;

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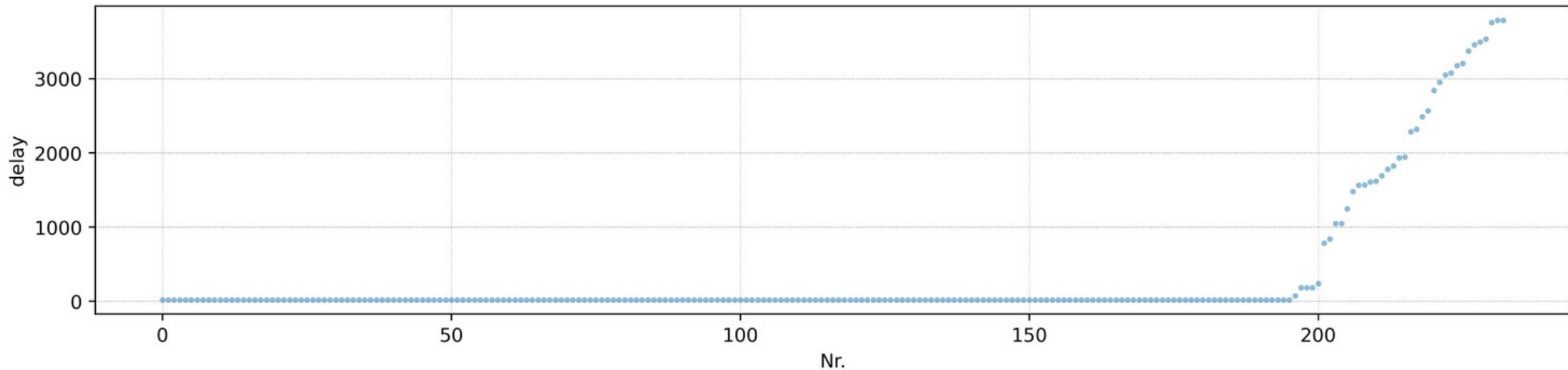
Timestamp: 1627590284 (29.07.2021. 20:24:44)

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

Different Links
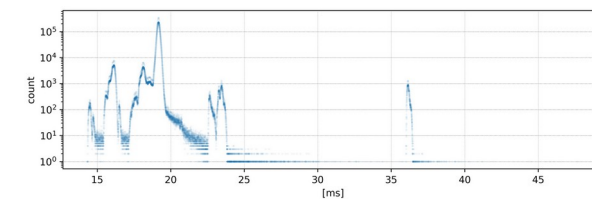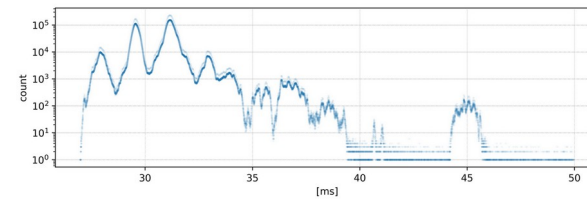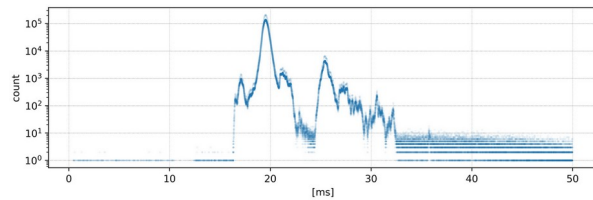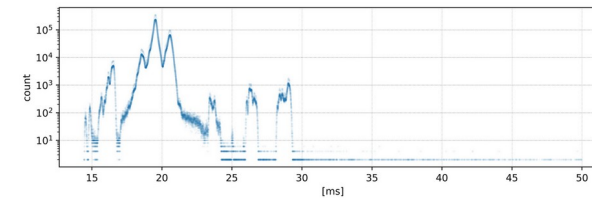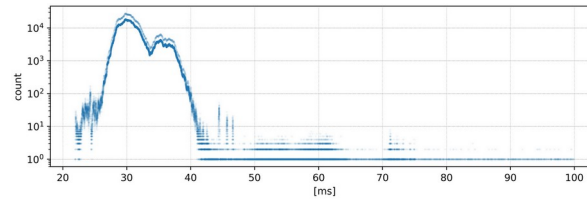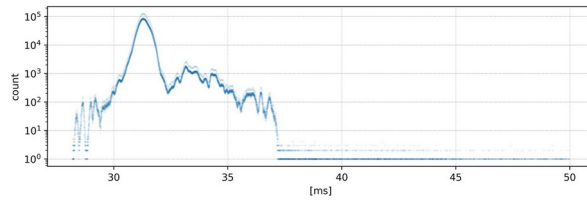
Measurement Errors

www.geant.org
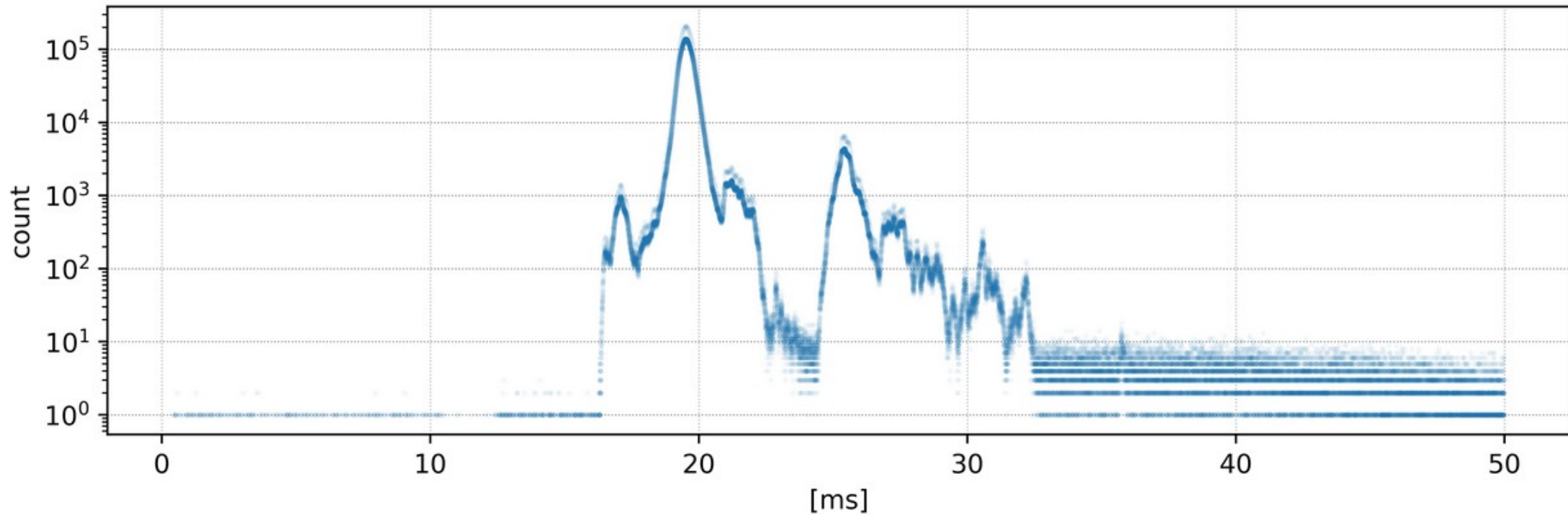
OWD: Same Endpoints - Different Directions
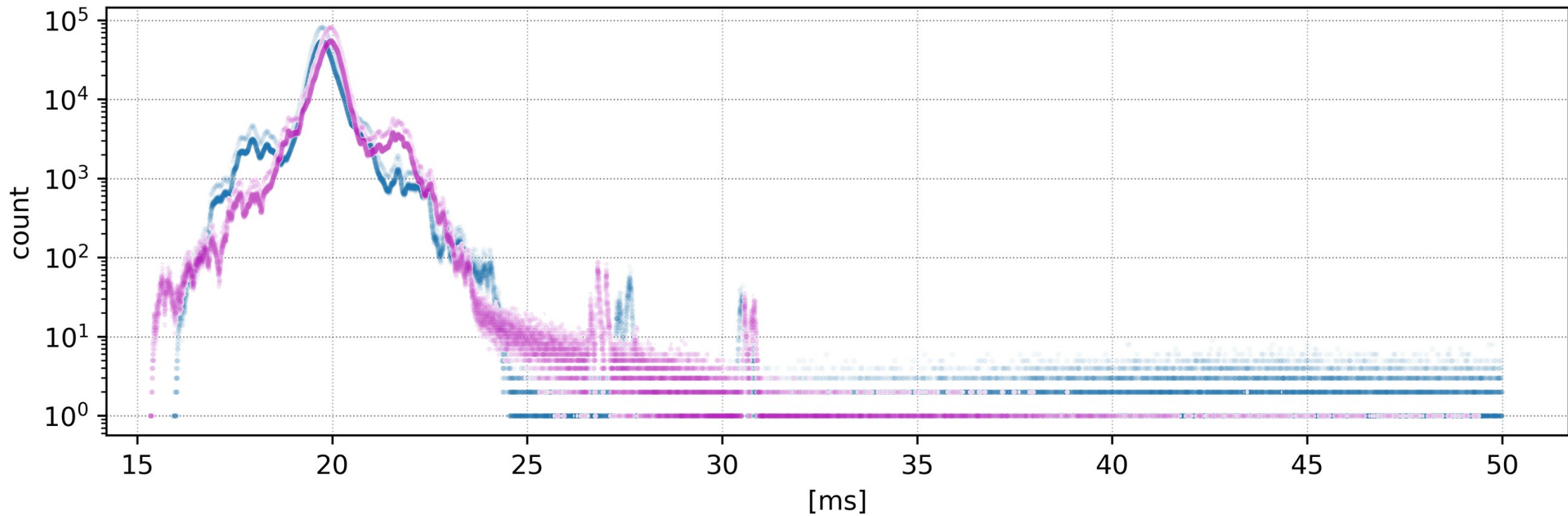
Throughput – Delay – Jitter Outlier (No Correlation)

www.geant.org
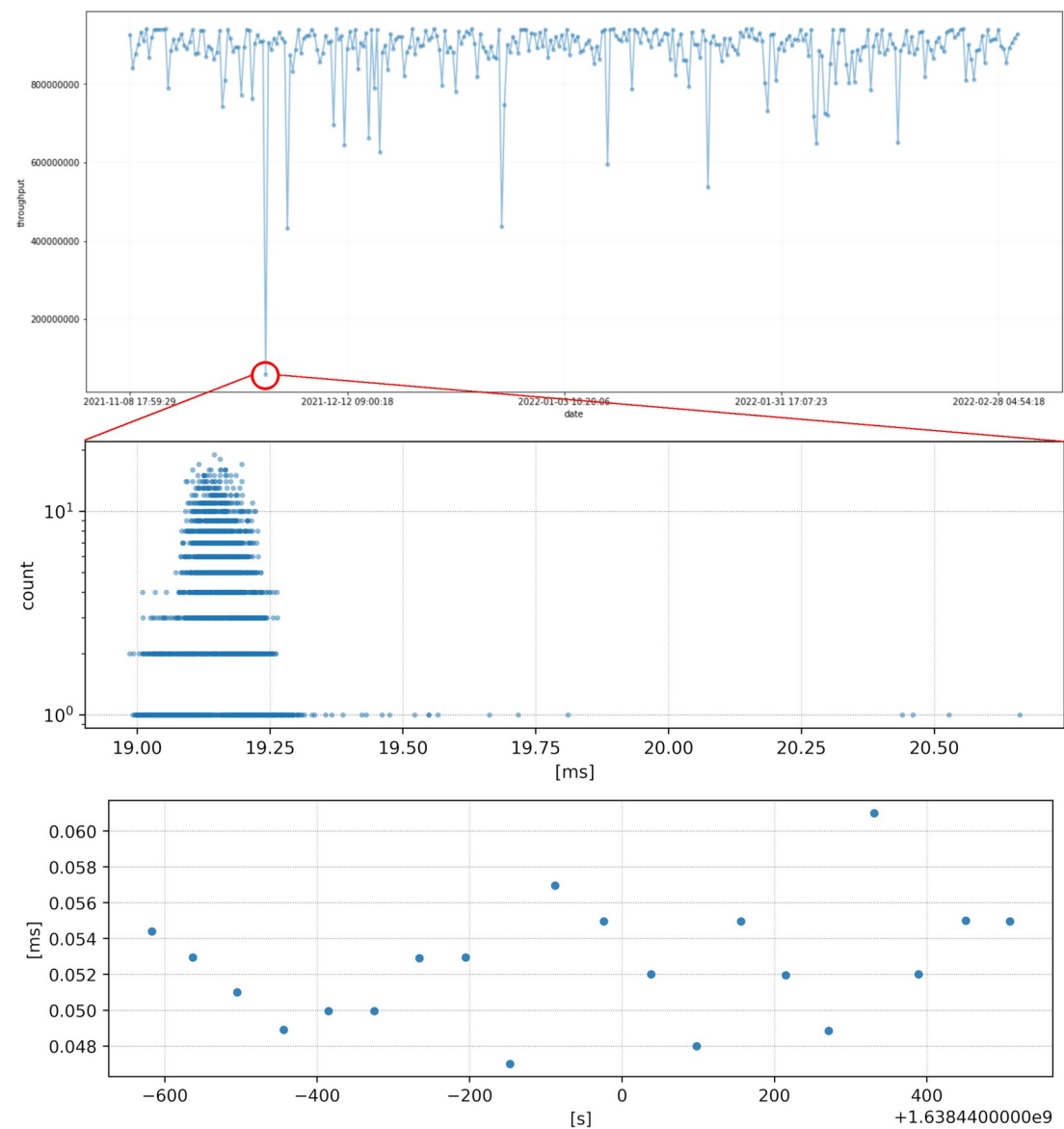
Throughput – Delay – Jitter Outlier (Correlation)

www.geant.org
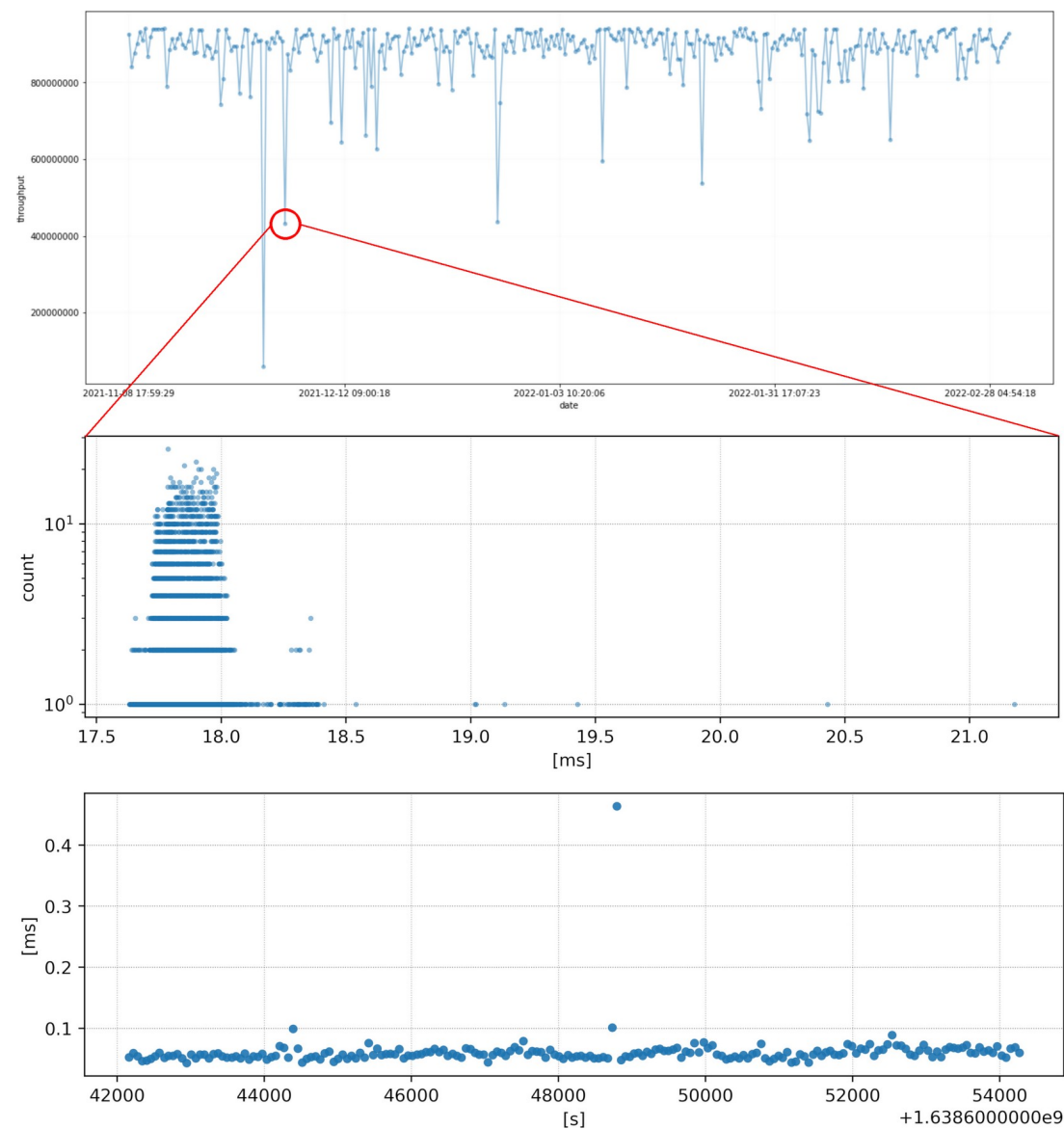
Current and Future Work

- 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

www.geant.org

Thank you

Any questions?

ljubomir@carnet.hr

www.geant.org