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/
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
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
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
  - ...

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


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
Thank you

Any questions?
ljubomir@carnet.hr

www.geant.org