



Network Optimized Transfers for Experimental Data (NOTED)

XrootD and FTS workshop - Ljubljana

28th Mars 2023

Carmen Misa Moreira and Edoardo Martelli

NOTED

Network Optimized Transfers of Experimental Data

History

From an idea discussed at the LHCCONE meeting #37 at BNL in 2017

Project started by CERN in 2018, funded by WLCG

Supervision: Edoardo Martelli and Tony Cass

Real work:

- Coralie Busse-Grawitz (2019)
- Joanna Waczynska (2020)
- Carmen Misa Moreira (2021-today)

Rationale

Problem:

- Large HEP data transfers can easily saturate WAN network links
- Routing protocols don't take into account link utilization and may force traffic to go over already congested paths, while alternative paths may be left idle

Goal:

- Reduce file transfer duration
- Optimize utilization of networks avoiding congestion and reducing idle time

Challenges:

- How long a large transfer will last? Is it worth to change the network now?
- When a corrective action should be taken? How long for?

How to approach the problem

Possible solutions:

- A) The application originating the large flows informs the network
- B) The Network tries to predict the behaviour of the application

Pros and Cons

Solution **A) the Application originating the large flows informs the network**

+ more precise: the application knows better what it is going to do

- more dependencies: requires commitment and involvement of developers and service transfers managers, definition of interfaces...

Solution **B) the Network tries to predict the behaviour of the application**

+ less invasive: no need to modify transfer applications

+ more effective: the network knows better where to improve itself

- less precise: traffic predictions may not be very accurate

NOTED approach

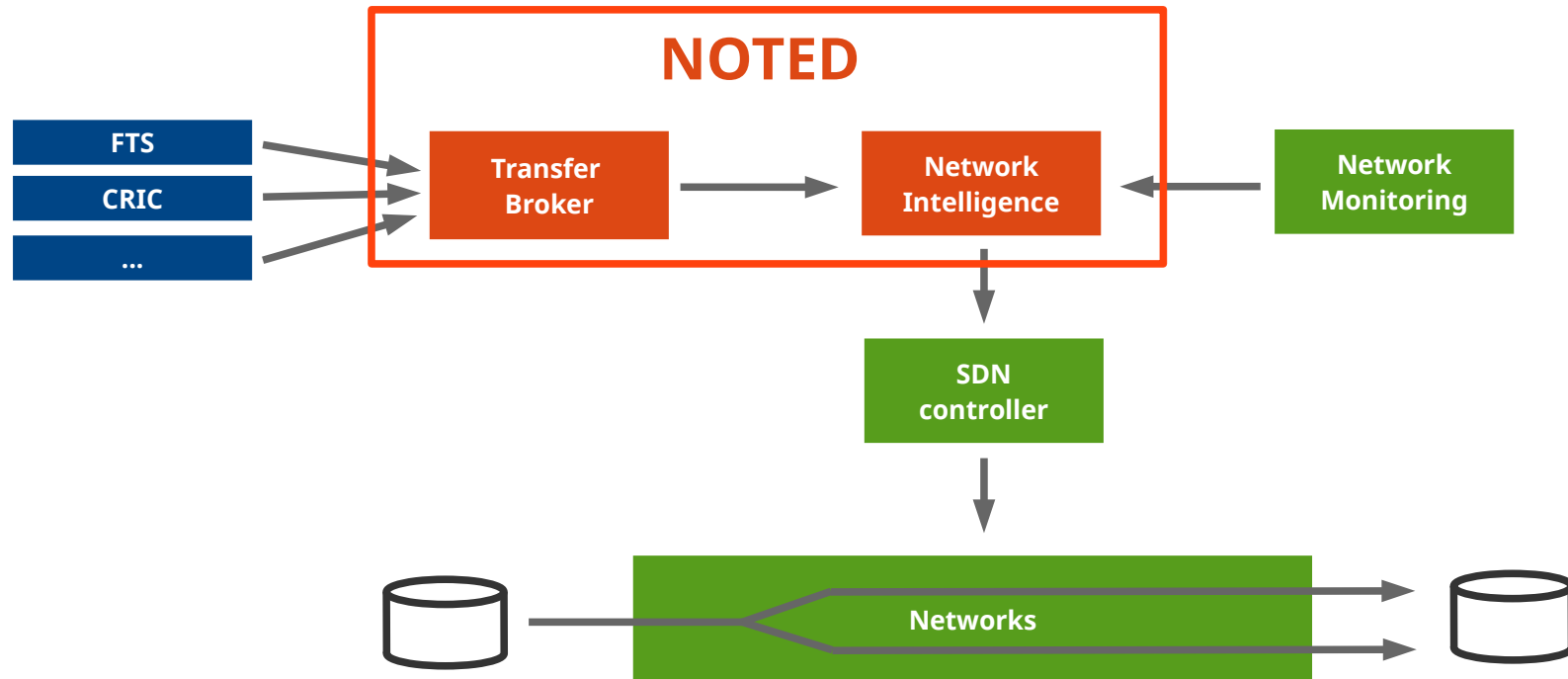
B) the Network tries to predict the behaviour of the applications

NOTED is an intelligent network controller that:

- continuously ***checks the status and the logs of the applications*** that generate large data transfers
- ***knows the links that may get congested*** and the source-destination sites that can originate the congesting transfers
- ***tries to predict the duration of large transfers***
- ***takes corrective actions only if the transfers are long lived***

Architecture

NOTED: framework that dynamically improves network performances for **large, on-going, long-lasting** data transfers



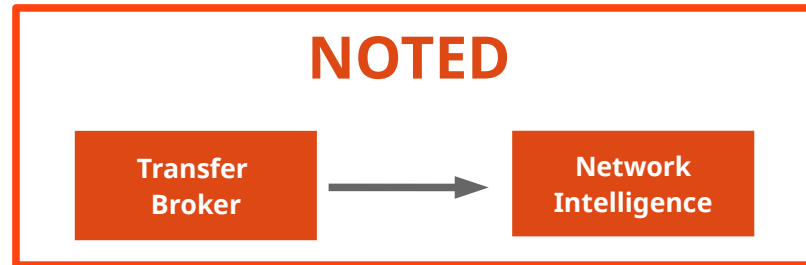
Components

Transfer Broker

The interface with the data transfer applications

Network Intelligence

It estimates the impact on the network and takes corrective actions

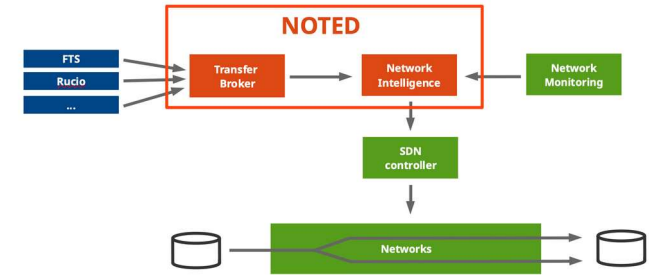


Transfer Broker

Transfer Broker (TB)

The interface with the transfer applications

- The operator configure the TB with the names of the sites that may congest the network with their transfers
- the TB may also queries resource databases to extract network information about the sites involved in the relevant transfers
- the TB queries the file transfer applications to know about on-going transfers and to get relevant metrics to determine their duration. The applications don't need to be aware of the work of NOTED

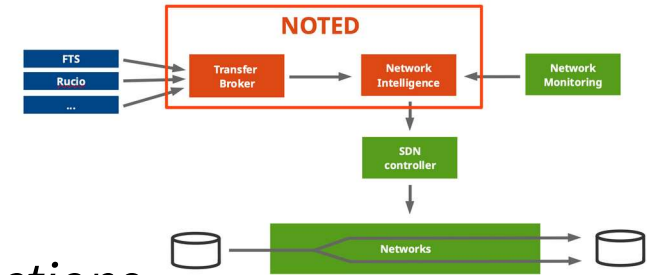


Network Intelligence

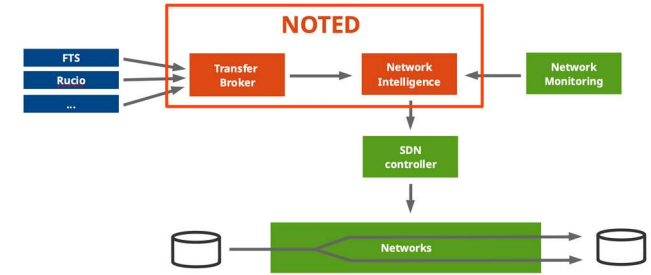
Network Intelligence (NI)

estimates the impact on the network, takes corrective actions

- Based on the information collected by the TB, the NI estimates the bandwidth that will be needed and the duration of the transfers
- IF the needed bandwidth could congest the network AND the transfer will last more than a minimum interval, THEN the NI triggers traffic engineer actions in the network
- Once the transfer is completed, the NI removes the changes it requested



Data Transfer applications

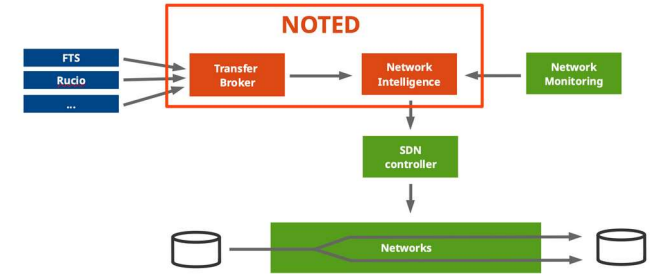


FTS *File Transfer Service*

- The current NOTED implementation works only with `FTS`
- The TB retrieves FTS metrics from the CERN MONIT Infrastructure
- Relevant parameters collected:
 - **{source se, dest se}**: source and destination endpoints involved in the transfer
 - **{throughput, filesize avg}**: throughput [bytes/s] and filesize [bytes] of the transfer
 - **{active count, success rate}**: number of TCP parallel flows and successful rate of the transfer
 - **{submitted count, connections}**: number of transfers in the queue and maximum number of transfers that can be held
- The TB enriches the FTS transfers with the IP prefixes of the involved sites found in the CRIC (Computer Resource Information Catalogue) database
- For future work: NOTED to modify the parameter of the FTS optimizer to increase the throughput

SDN controllers

- **AutoGOLE - SENSE**
- **Any SDN**

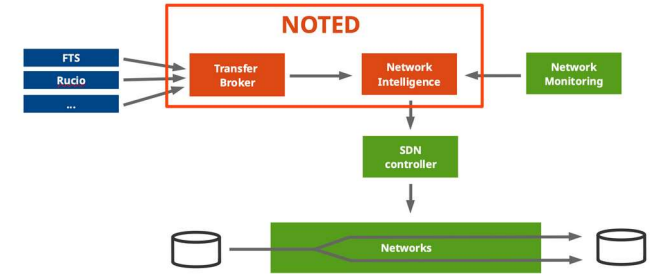


The NI can use the API provided by SDN controllers to take corrective actions in the network

So far it has been tested with:

- SENSE to request direct circuits between the source and destination sites
- Juniper APIs to re-configure BGP metrics and load balance traffic over multiple links

Network monitoring



Before taking the corrective actions, the NI should make sure that the network path where it wants to offload the transfer is available and not already overloaded

For this, the decision algorithm should take into account real time links utilization statistics coming from network monitoring systems

For future work

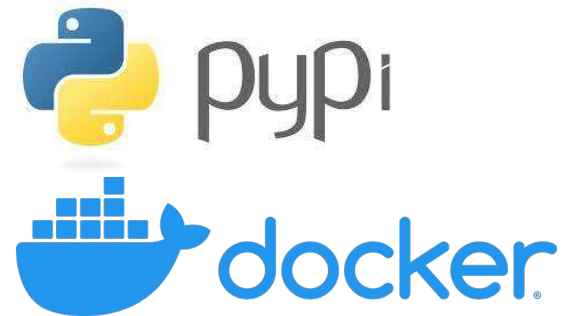
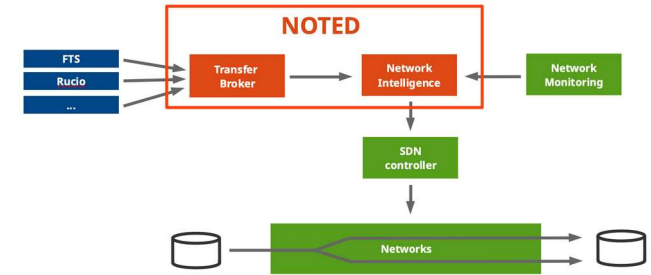
NOTED status

Version 2 already released:

- rewritten in Python
- improved efficiency and stability
- easier configuration
- open source (GPL v3)

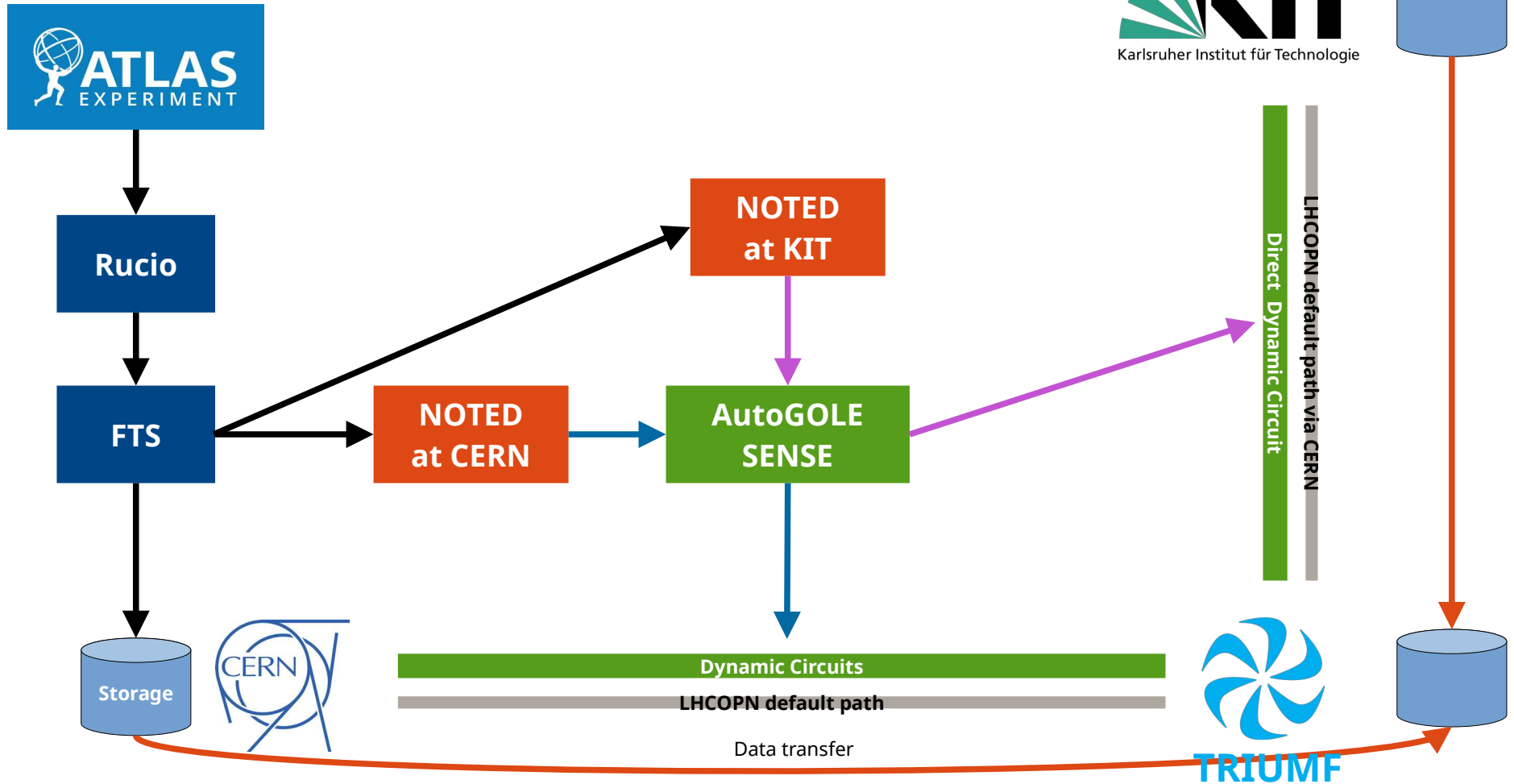
Package distribution:

- available at <https://pypi.org/project/noted-dev/>
- also as docker container:
<https://hub.docker.com/r/carmenmisa/noted-docker>



NOTED trials

NOTED demo for SC22



NOTED demo

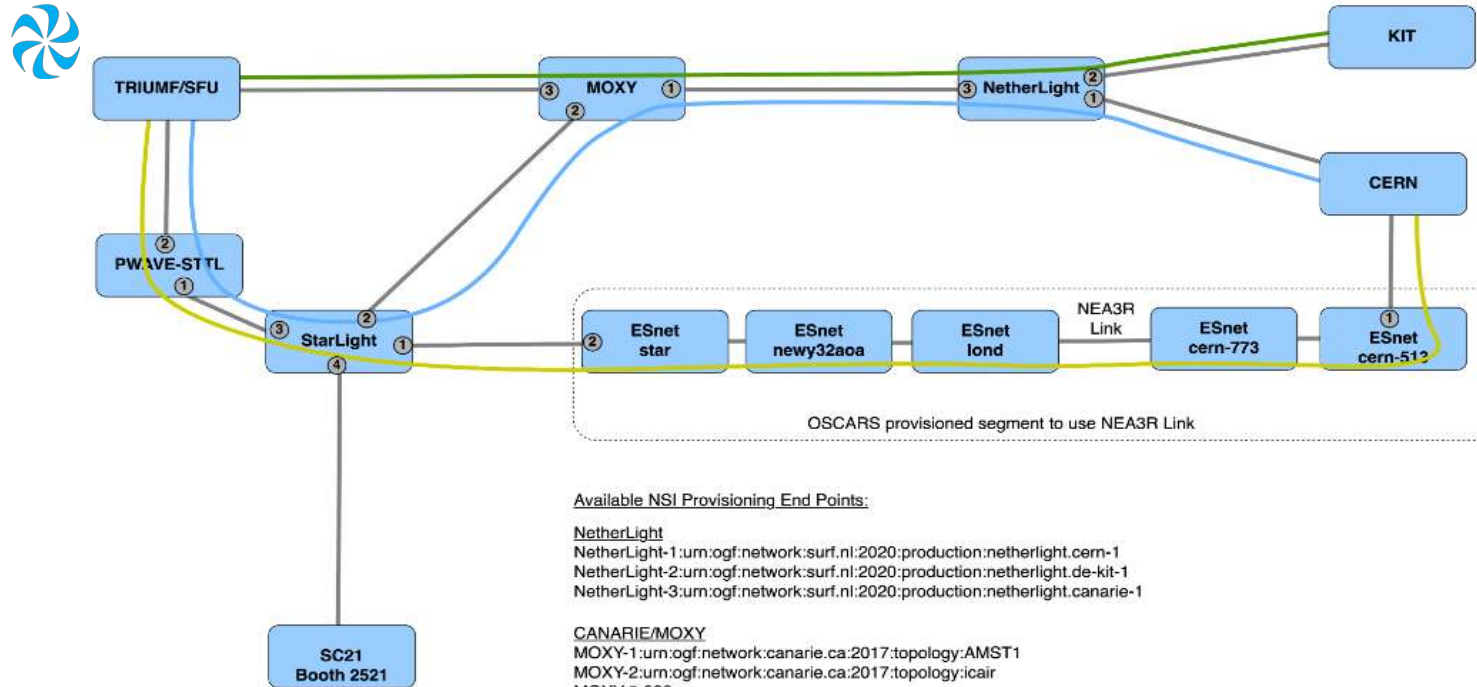
The NOTED controller looks for large FTS transfers from CH-CERN to CA-TRIUMF and from DE-KIT to CA-TRIUMF

Once a large transfers is detected, NOTED requests a direct circuit CERN-TRIUMF or KIT-TRIUMF to the AutoGOLE/SENSE provisioning system

The routers at CERN, KIT, TRIUMF have BGP peerings pre-configured on the dynamic circuits with better metrics. When the circuits are activated, the traffic get re-routed over them

When a large transfer is completed, the dynamic circuit is released and the traffic is routed back to the production LHCOPN links

Dynamic circuits



- Link #1 - VLAN 2025
- Link #2 - VLAN 2024
- Link #3 - VLAN 3694

Available NSI Provisioning End Points:

NetherLight

NetherLight-1:urn:ogf:network:surf.nl:2020:production:netherlight.cern-1
 NetherLight-2:urn:ogf:network:surf.nl:2020:production:netherlight.de-kit-1
 NetherLight-3:urn:ogf:network:surf.nl:2020:production:netherlight.canarie-1

CANARIE/MOXY

MOXY-1:urn:ogf:network:canarie.ca:2017:topology:AMST1
 MOXY-2:urn:ogf:network:canarie.ca:2017:topology:icair
 MOXY-3:???

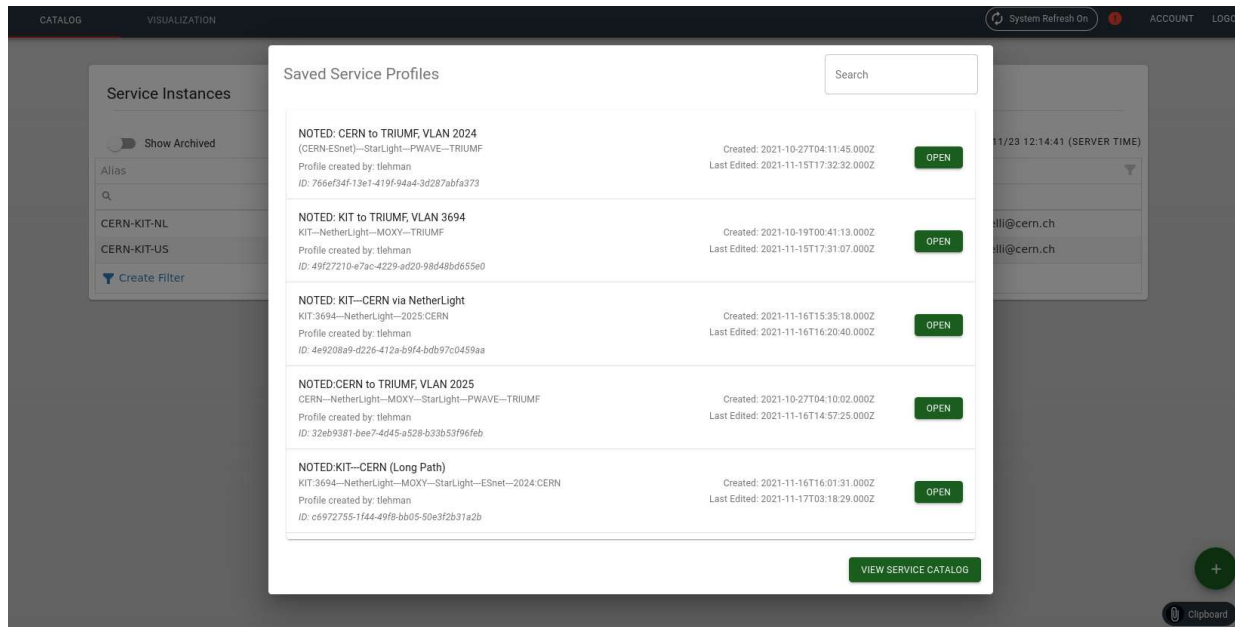
ESnet

ESnet-1:urn:ogf:network:es.net:2013::cern-513-cr5:lag-2:+
 ESnet-2:urn:ogf:network:es.net:2013::star-cr55:2_1_c3_1:+

StarLight

StarLight-1:urn:ogf:network:icair.org:2013:mren8700:esnet
 StarLight-2:urn:ogf:network:icair.org:2013:mren8700:canarie
 StarLight-3:urn:ogf:network:icair.org:2013:mren8700:pwave-grp
 StarLight-4:???

ESnet SENSE provisioning system



The screenshot displays the SENSE web UI interface. A modal window titled "Saved Service Profiles" is open, showing a list of profiles. Each profile entry includes a title, a description, the creator's name, creation and last edit timestamps, and an "OPEN" button. The profiles listed are:

- NOTED: CERN to TRIUMF, VLAN 2024** (CERN-ESnet)–StarLight–PWave–TRIUMF. Profile created by: tlehman. ID: 766ef34f-13e1-419f-94a4-3d287abfa373. Created: 2021-10-27T04:11:45.000Z. Last Edited: 2021-11-15T17:32:32.000Z.
- NOTED: KIT to TRIUMF, VLAN 3694** KIT–NetherLight–MOXY–TRIUMF. Profile created by: tlehman. ID: 49f27210-e7ac-4229-ad20-98d48bd655e0. Created: 2021-10-19T00:41:13.000Z. Last Edited: 2021-11-15T17:31:07.000Z.
- NOTED: KIT–CERN via NetherLight** KIT:3694–NetherLight–2025.CERN. Profile created by: tlehman. ID: 4e9208a9-d226-412a-b9f4-bbb97c0459aa. Created: 2021-11-16T15:35:18.000Z. Last Edited: 2021-11-16T16:20:40.000Z.
- NOTED:CERN to TRIUMF, VLAN 2025** CERN–NetherLight–MOXY–StarLight–PWave–TRIUMF. Profile created by: tlehman. ID: 32eb9381-bee7-4d45-a52b-b33b53f9f6eb. Created: 2021-10-27T04:10:02.000Z. Last Edited: 2021-11-16T14:57:25.000Z.
- NOTED:KIT–CERN (Long Path)** KIT:3694–NetherLight–MOXY–StarLight–ESnet–2024.CERN. Profile created by: tlehman. ID: c6972755-1f44-49f8-bb05-50e3f2b31a2b. Created: 2021-11-16T16:01:31.000Z. Last Edited: 2021-11-17T03:18:29.000Z.

At the bottom of the modal, there is a "VIEW SERVICE CATALOG" button. The background shows a sidebar with "Service Instances" and "Show Archived" options, and a main content area with a search bar and a list of instances.

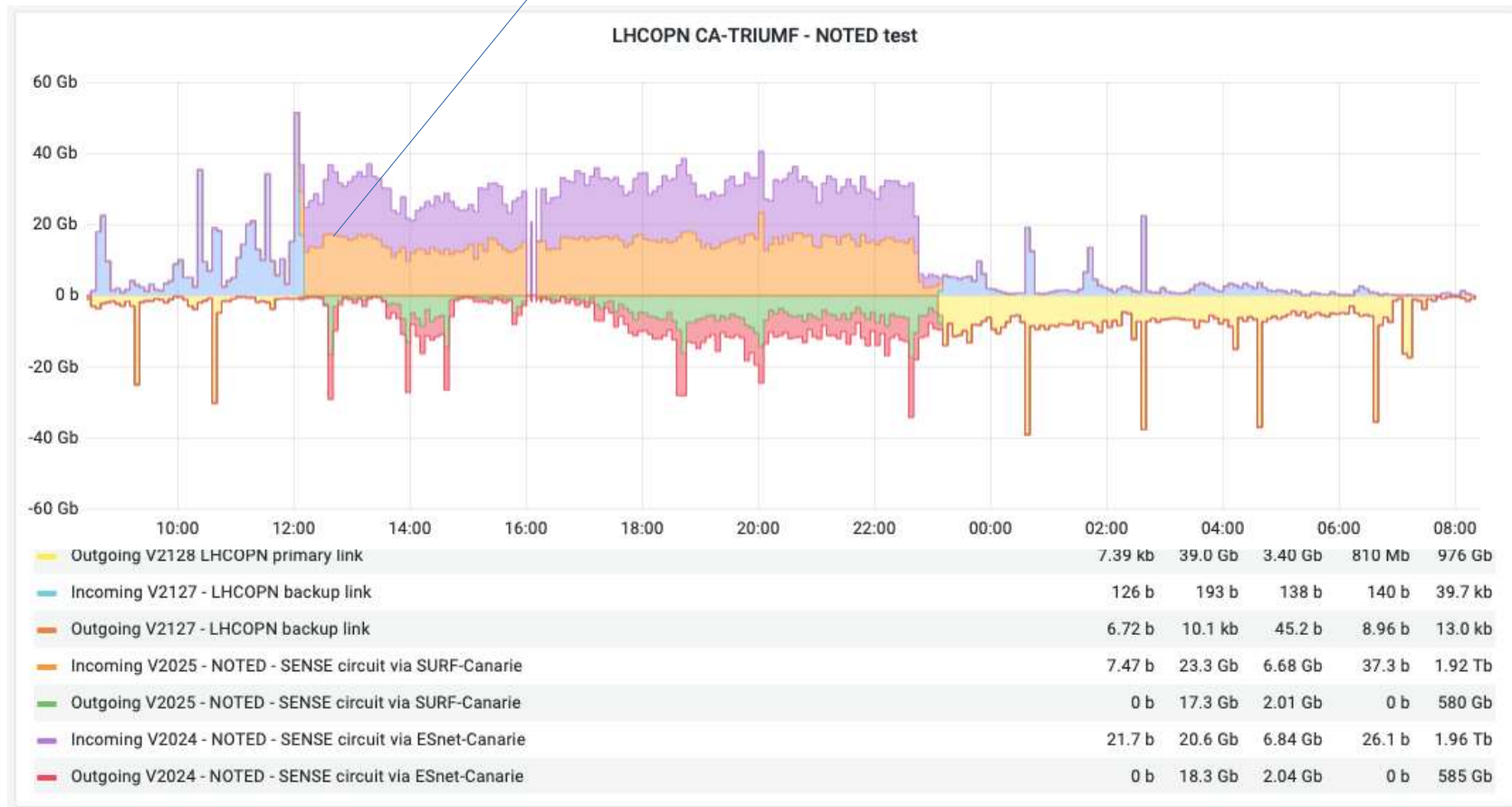
SENSE web UI

```
#!/bin/sh
UUID="57041c58-0f4f-45dd-b5be-20ca5624d430"
STATUS=`/usr/local/bin/sense_util.py -s -u ${UUID}`
echo "Provision request from NOTED"
if [ "$STATUS" == "CANCEL - READY" ] ; then
    /usr/local/bin/sense_util.py -r -u ${UUID}
fi
```

SENSE API

SC22 demo

✳️ NOTED activated circuits



Machine Learning

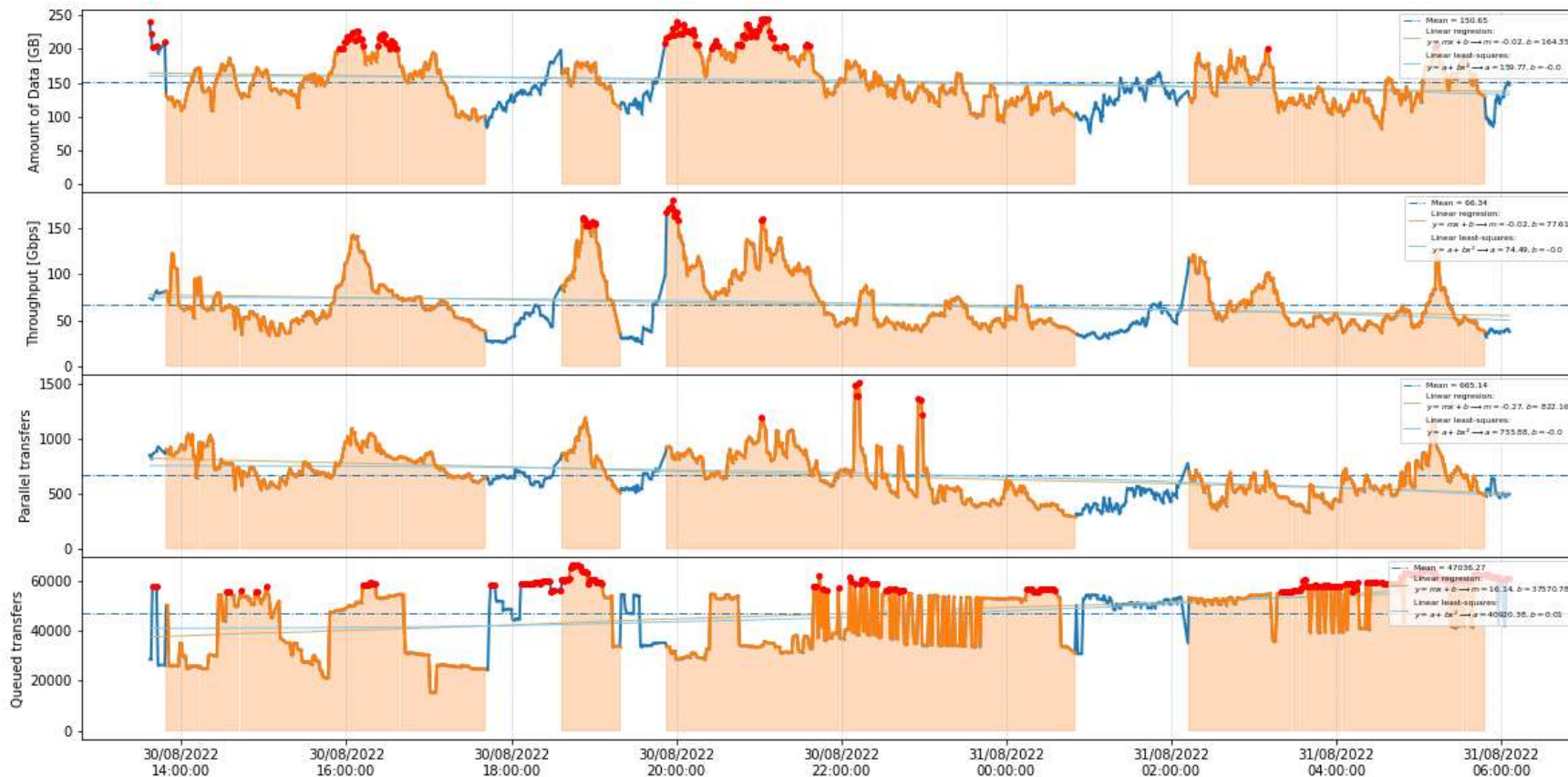
Machine learning

Machine Learning LSTM has been tested to better estimate the duration and the volume of the FTS transfers

Work in progress

“Plain” NOTED in actions

LHCONE 31th of August 2022

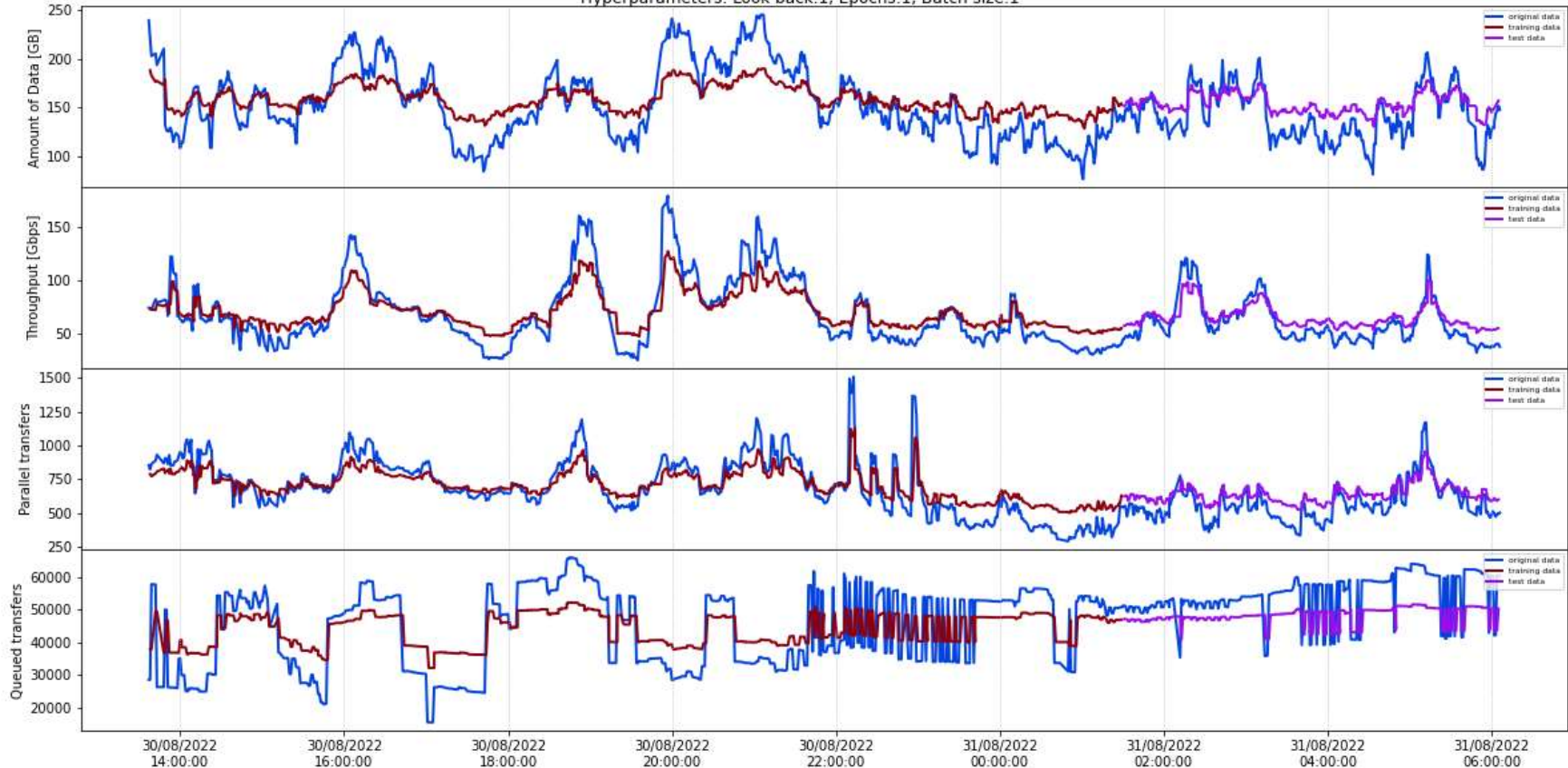


Orange area: NOTED triggered network action

Traffic forecast with LSTM

Long-Short Term Memory Machine Learning Algorithm
Traffic Forecasting
LHCONE 31th of August 2022

Hyperparameters: Look back:1, Epochs:1, Batch size:1



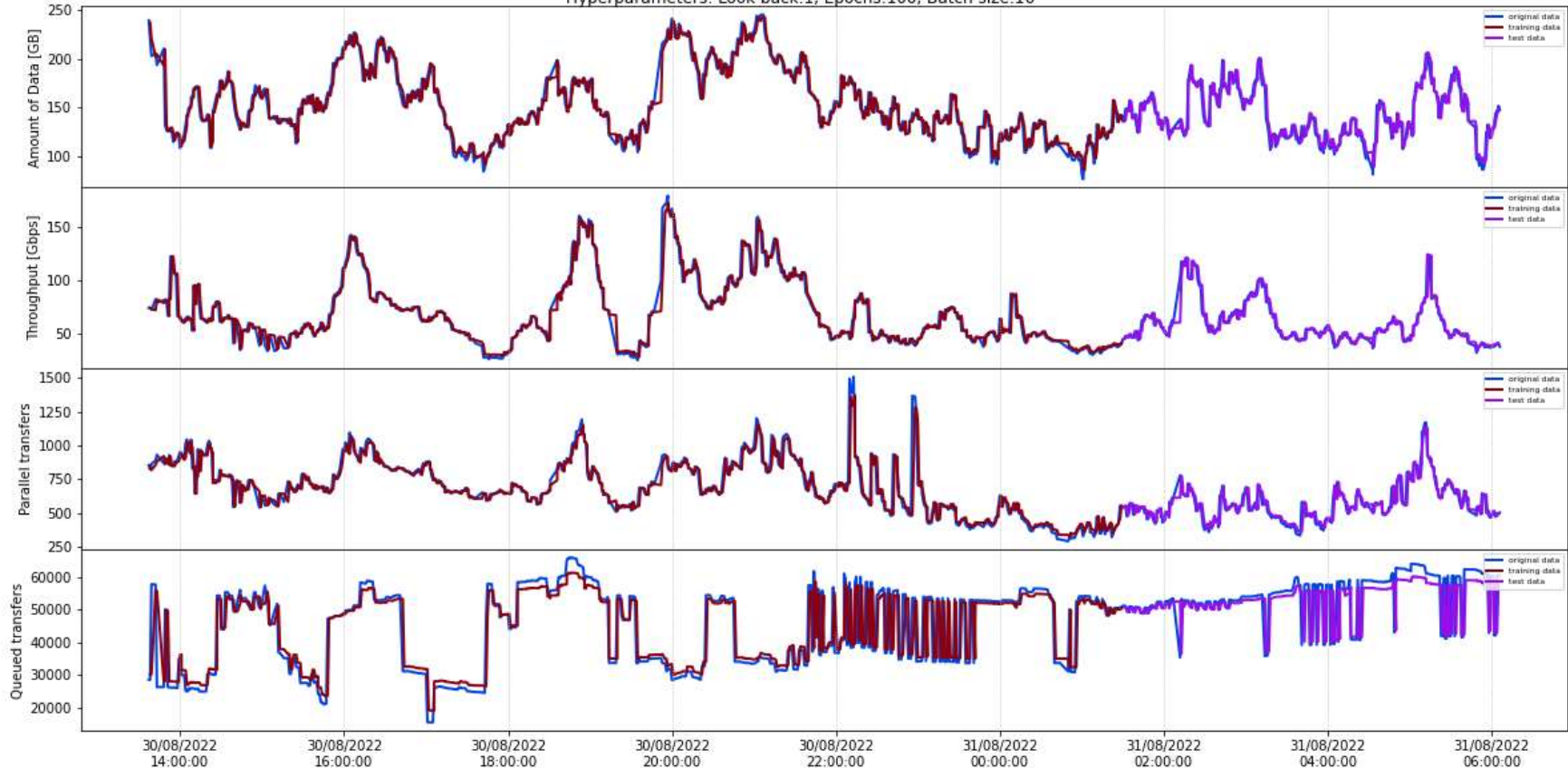
This model is not well fitted to the data

Real data
Training data
Predicted data

Traffic forecast with LSTM

Long-Short Term Memory Machine Learning Algorithm
Traffic Forecasting
LHCONE 31th of August 2022

Hyperparameters: Look back:1, Epochs:100, Batch size:16



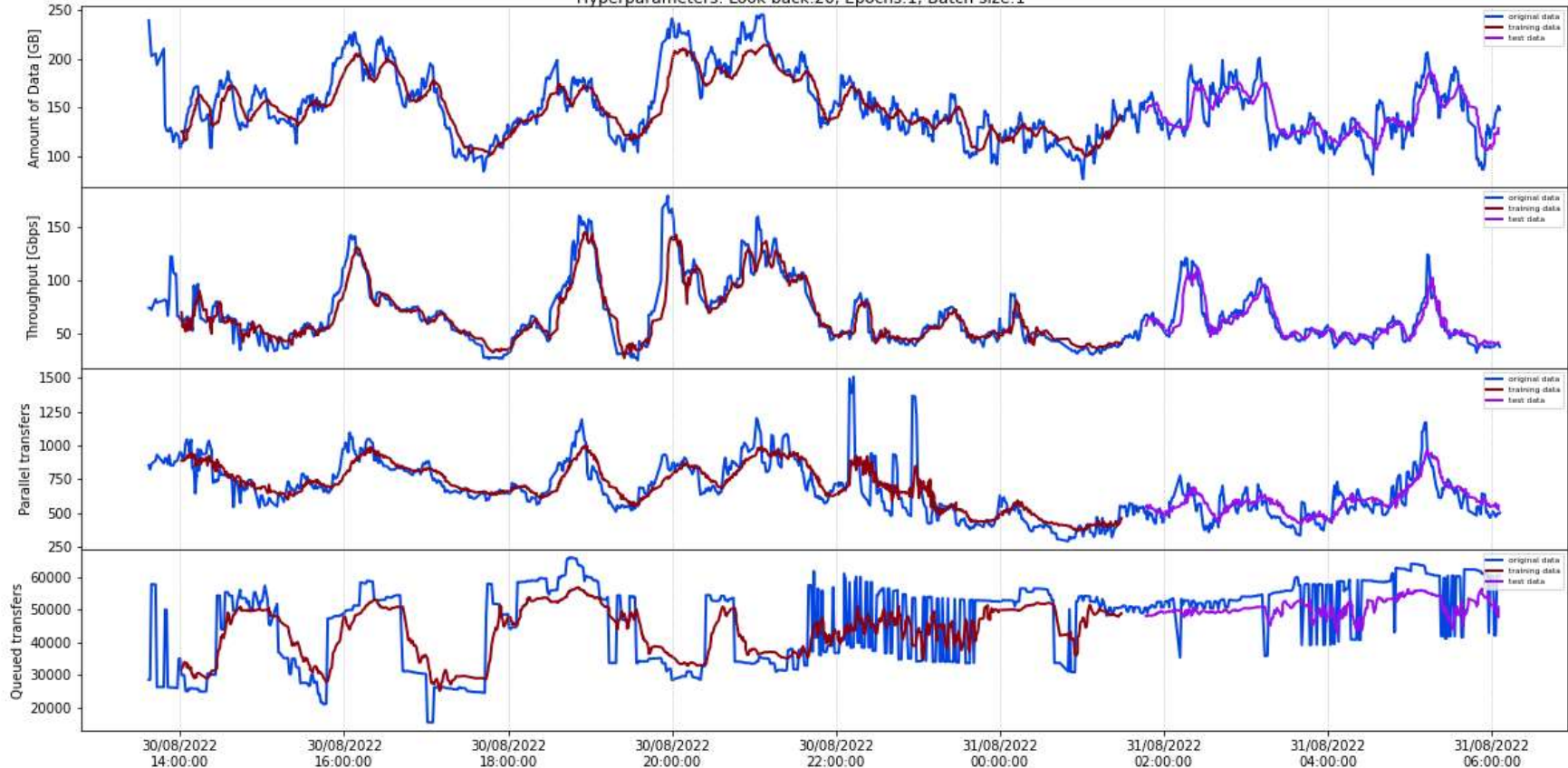
With increased Epoch and Batch size, the model fits very well

Real data
Training data
Predicted data

Traffic forecast with LSTM

Long-Short Term Memory Machine Learning Algorithm
Traffic Forecasting
LHCONE 31th of August 2022

Hyperparameters: Look back:20, Epochs:1, Batch size:1



With increased loopback (20), the model fits well even with epoch 1 and batch 1

Real data
Training data
Predicted data

Execution details

Look back: 1 Epochs: 1 Batch size: 1

CPU times: user 3.45 s, sys: 120 ms, **total: 3.57 s**

Peak memory: 711.70 MiB

Train Score: 15.77 RMSE

Test Score: 11.37 RMSE

Length of train dataset: 821

Length of test dataset: 353

Look back: 1 Epochs: 100 Batch size: 16

CPU times: user 16.3 s, sys: 578 ms, **total: 16.8 s**

Peak memory: 816.57 MiB

Train Score: 6.96 RMSE

Test Score: 5.59 RMSE

Length of train dataset: 821

Length of test dataset: 353

Look back: 20 Epochs: 1 Batch size: 1

CPU times: user 5.57 s, sys: 138 ms, **total: 5.7 s**

Peak memory: 823.27 MiB

Train Score: 11.83 RMSE

Test Score: 9.22 RMSE

Length of train dataset: 821

Length of test dataset: 353

Future research

Use autoencoders and transformers

Make predictions in real time

Conclusions

Conclusions

NOTED can reduce data transfers duration and improve the efficient use of network resources

NOTED makes decision by watching and understanding the behaviour of transfer services. Transfer Applications don't need any modification to work with NOTED

NOTED capabilities have been demonstrated with production FTS transfers

Resources

- NOTED project
- NOTED software

Questions?

edoardo.martelli@cern.ch
carmen.misa@cern.ch

