CERN

Diving into large-scale congestion with NOTED as a network controller and machine learning traffic forecasting

CERN
IT Department CS Group NE Section

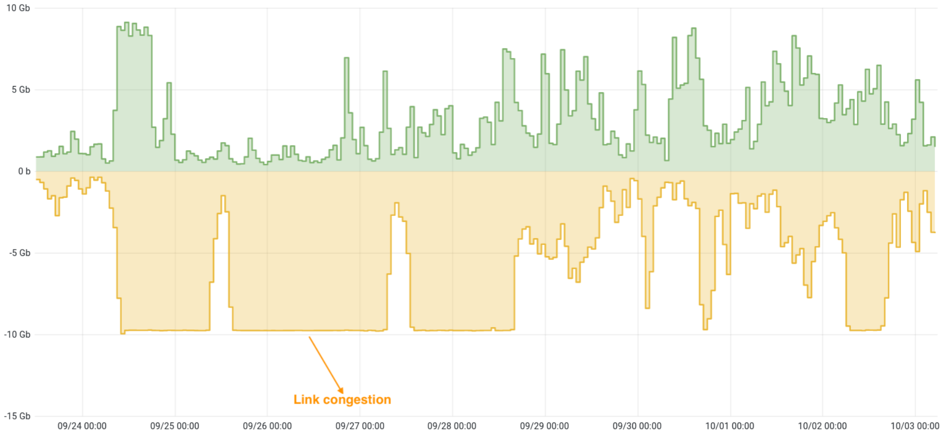
27th International Conference on Computing in High Energy & Nuclear Physics (CHEP24)

Carmen Misa Moreira (CERN)
Elisabetta Schneider (ETH Zürich, summer student 2024)
Edoardo Martelli (CERN)

Outline

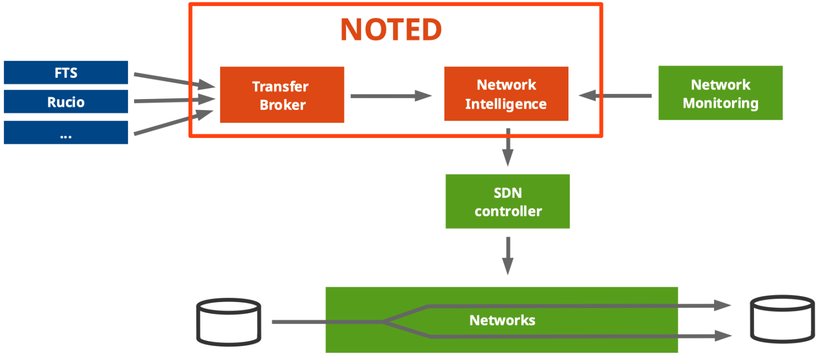
Introduction

Motivation



- Large data transfers can saturate network links while alternative paths may be left idle

Architecture



NOTED (Network Optimized Transfer of Experimental Data)

An intelligent network controller to improve the throughput of large data transfers in FTS (File Transfer Services) by handling dynamic circuits or by doing load balance.

Elements

FTS (File Transfer Service):

- Analyse data transfers to estimate if any action can be applied to optimise the network utilization → get on-going and queued transfers.

CRIC (Computing Resource Information Catalog):

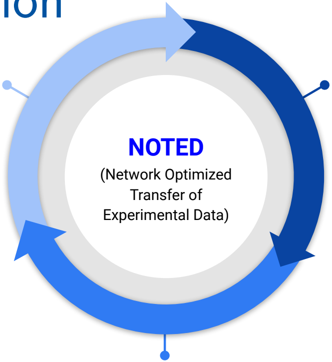
- Use the CRIC database to get an overview of the network topology → get IPv4/IPv6 addresses, endpoints, rcsite and federation.

FTS File Transfer Service

CRIC Computing Resource Information Catalog

elasticsearch

Modes of operation



CUSTOM

NOTED is working based on the parameters written in a config.yaml file by the network administrator to monitor FTS data transfers

LHCOPN

When CERN NMS raises an alarm on an interface in one of the LHCOPN border routers, NOTED identifies the Tier 1 and starts to monitor FTS data transfers → automatically!

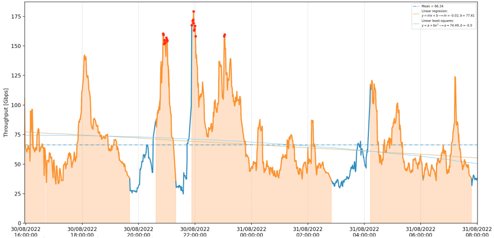
LHCONE

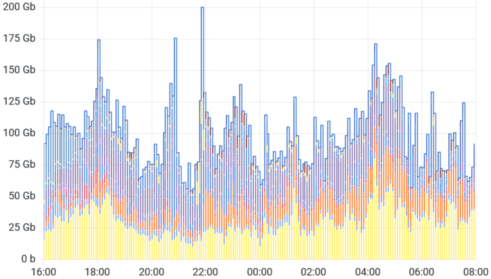
When CERN NMS raises an alarm on an interface in one of the LHCONE border routers, NOTED identifies the Tier 2, Tier 3 and starts to monitor FTS data transfers → automatically!

- Much more complex for LHCONE since a single path is shared by multiple sites $\sim$ 100.

NOTED demonstrations

Transfers of WLCG sites in LHCONE (31st of August 2022)



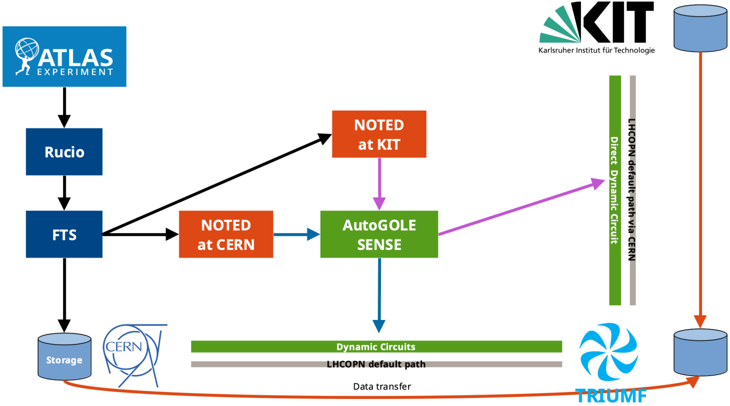


- If throughput $>$ 80 GB/s $\rightarrow$ NOTED provides a dynamic circuit. When throughput $<$ 40 GB/s $\rightarrow$ NOTED cancels the dynamic circuit and the traffic is routed back to the default path.

- Observations of NOTED about the network utilization correspond with the reported ones in Grafana by LHCONE/LHCOPN production routers.

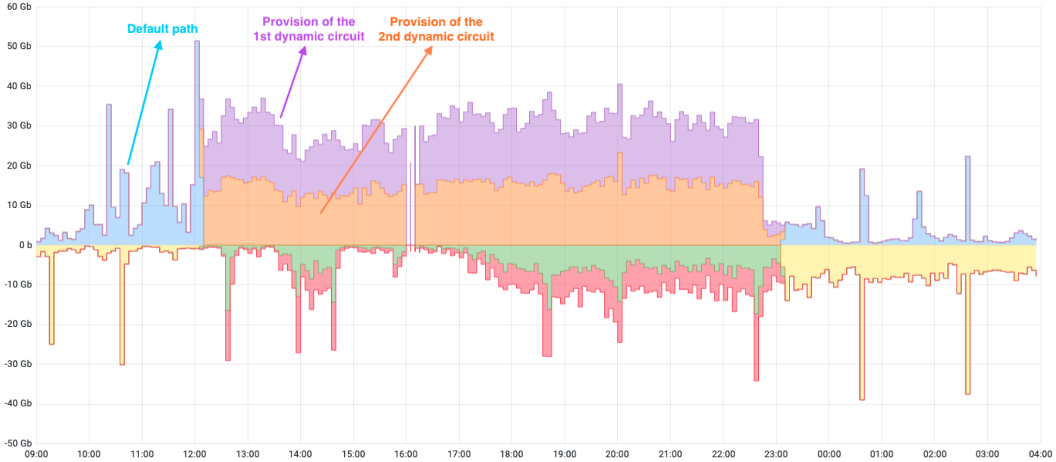
Therefore, by inspecting FTS data transfers it is possible to get an understanding of the network usage and improve its performance by executing an action in the topology of the network.

NOTED demo at SC22 (CUSTOM version)

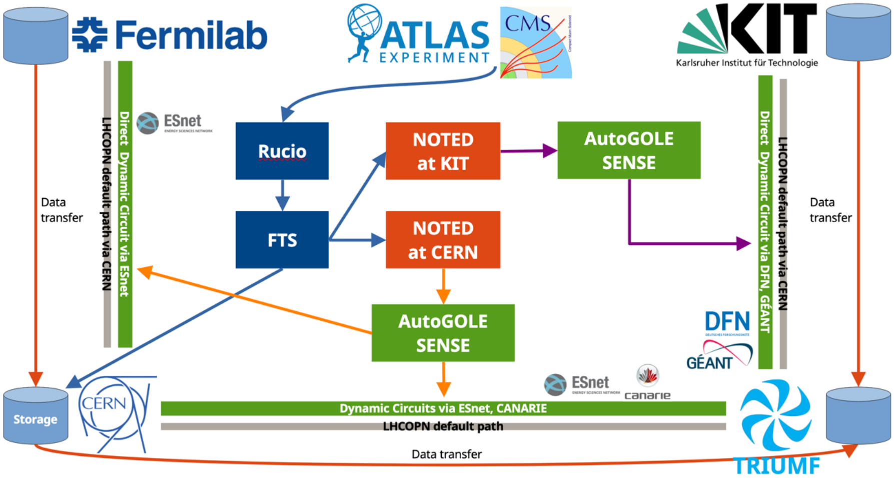


1. NOTED looks in FTS for large data transfers.
2. When it detects a large data transfer → request a dynamic circuit by using the SENSE/AutoGOLE provisioning system.
3. LHCOPN routers at CERN will route the data transfers over the new dynamic circuit.
4. When the large data transfer is completed → release the dynamic circuit, the traffic is routed back to the LHCOPN production link.

NOTED demo at SC22 (CUSTOM version)

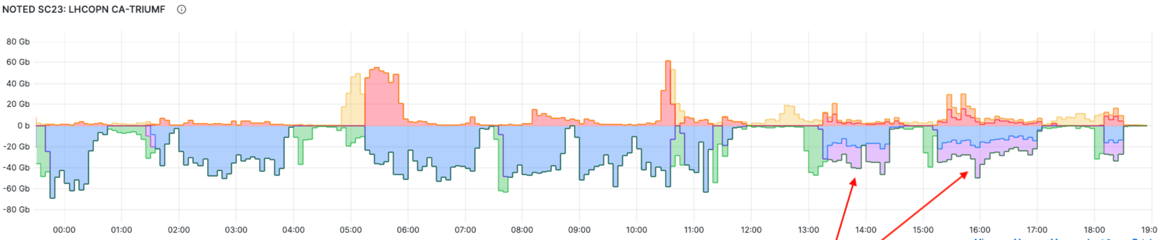


NOTED demo at SC23 (LHCOPN, LHCONE and custom versions)



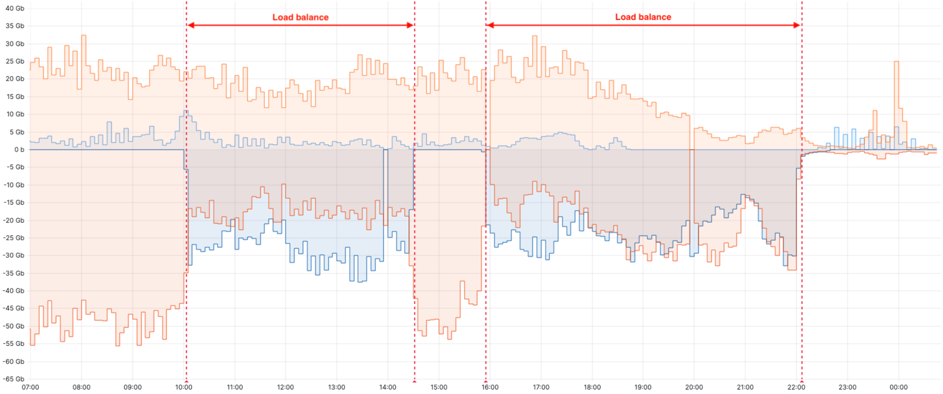
NOTED demo at SC23 (LHCOPN, LHCONE and custom versions)

- Results of 14th November 2023.
- Data transfers between CH-CERN - CA-TRIUMF through SC23 booth.



	Min	Max	Mean	Last *	Total
NOTED VLAN 2024 in - TRIUMF to CERN	0 b	61.7 Gb	4.13 Gb	0 b	966 Gb
NOTED VLAN 2024 out - CERN to TRIUMF	80 b	69.5 Gb	19.9 Gb	104 b	4.66 Tb
NOTED VLAN 2027 - SC23 Booth in - TRIUMF to CERN	0 b	14.1 Gb	710 Mb	0 b	164 Gb
NOTED VLAN 2027 - SC23 Booth out - CERN to TRIUMF	55.2 b	29.3 Gb	2.60 Gb	100 b	601 Gb
in total - TRIUMF to CERN	115 b	49.1 Gb	2.06 Gb	396 kb	483 Gb
out total - CERN to TRIUMF	151 b	54.5 Gb	3.98 Gb	189 kb	932 Gb

NOTED demo at DC24 (LHCOPN, LHCONE versions)

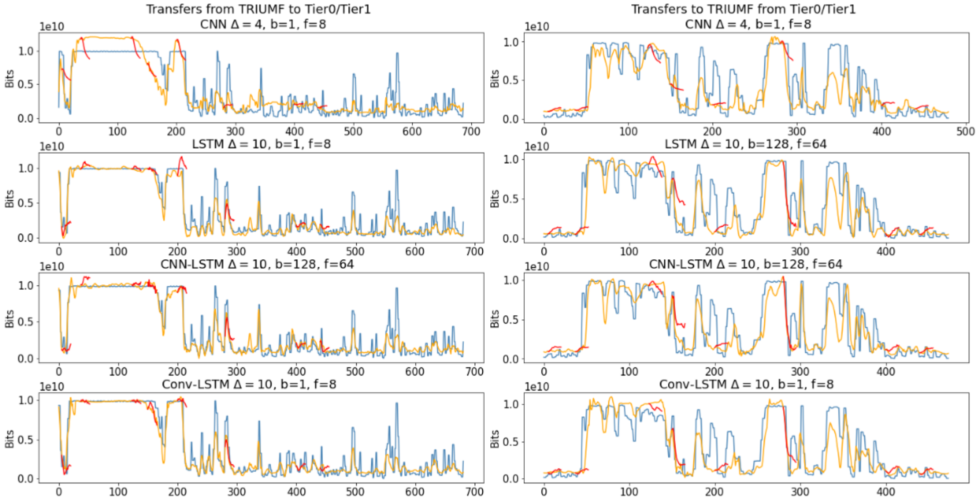


Name	Min	Max	Mean	Last *	Total
Backup Link Incoming to CERN	0 b	11.1 Gb	1.89 Gb	966 kb	406 Gb
Backup Link Outgoing from CERN	0 b	37.5 Gb	14.9 Gb	307 b	3.21 Tb
Primary Link Incoming CA-TRIUMF to CERN	0 b	32.4 Gb	15.5 Gb	437 Mb	3.33 Tb
Primary Link Outgoing CERN to CA-TRIUMF	0 b	55.6 Gb	24.3 Gb	962 Mb	5.22 Tb

- CA-TRIUMF load balancing between LHCOPN and its backup link (from 21st to 23rd of February 2024)

Machine learning network traffic forecasting

Initial study (presented at CHEP 2021)

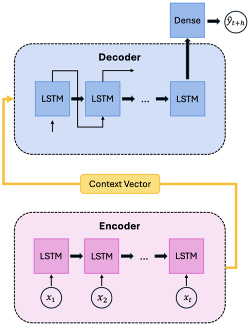


Initial study (presented at CHEP 2021)

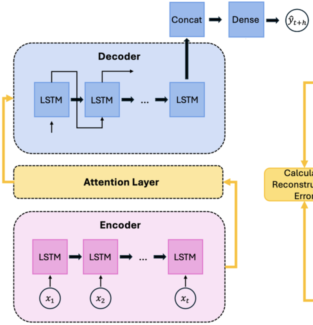
- The CNN layer helps to detect small differences → good for single-step forecast, not appropriate for long-term forecast.
- The LSTM is able to capture pattern and is not affecting by the window size.
- A CNN-LSTM combined model overcomes CNN limitations in terms of long temporal dependencies and achieves optimal forecast performance. It can better predict minor irregularities → more sensitive to short-term increases in traffic.
- The Conv-LSTM model is less sensitive to rapid changes, therefore, compared to CNN-LSTM, The Conv-LSTM is worse at detecting small transfers.

We consider CNN-LSTM as the best model for transfer prediction and Conv-LSTM as the most suitable model to predict the transfer end. Conv-LSTMs exhibit a slower reaction to drops, therefore, can prevent multiple reconfiguration of the network during a short period of time.

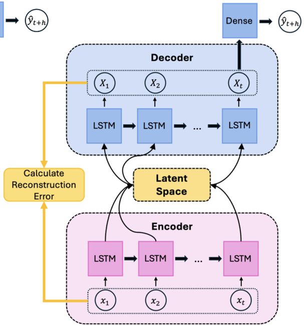
Network traffic forecasting models



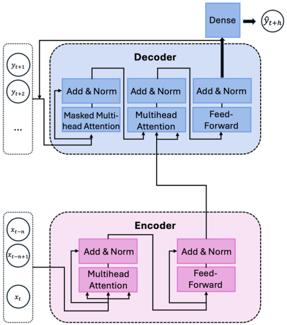
Seq2Seq encoder-decoder model



Seq2Seq encoder-decoder model with an attention layer



Autoencoder

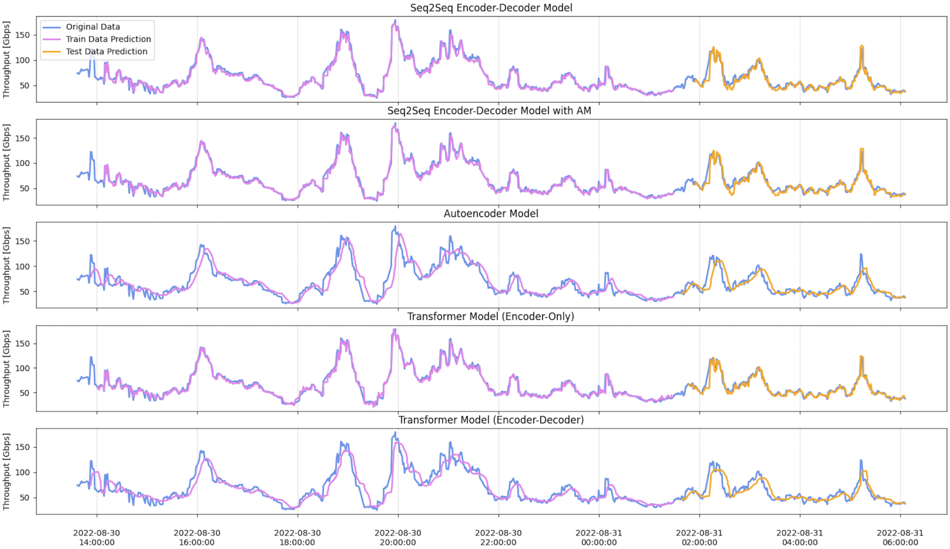


Encoder-decoder transformer

Network traffic forecasting models

- Seq2Seq: simpler, the encoder processes the input sequence and converts it into a single fixed-length context vector → struggles with long sequences.
- Seq2Seq with attention mechanism: allows the decoder to access different parts of the input sequence at each step rather than relying on the fixed-length context vector → handle long sequences but increases complexity and computational time.
- Autoencoder: encodes the input data into a lower-dimensional latent space, it can be trained with historical data → map past data to future predictions.
- Transformer: encoder with selft-attention mechanism, enables parallelization → speeds up training and inference, computationally expensive.

Results



Seq2Seq Encoder-Decoder Model
Seq2Seq Encoder-Decoder Model with AM
Autoencoder Model
Transformer Model (Encoder-Only)
Transformer Model (Encoder-Decoder)
Original Data
Train Data Prediction
Test Data Prediction
Throughput [Gbps]
150
100
50
2022-08-30 14:00:00
2022-08-30 16:00:00
2022-08-30 18:00:00
2022-08-30 20:00:00
2022-08-30 22:00:00
2022-08-31 00:00:00
2022-08-31 02:00:00
2022-08-31 04:00:00
2022-08-31 06:00:00

Results

Algorithm	RMSE	σ(RMSE)	r_s	CPU Time
Seq2Seq Encoder-Decoder model	5.783	0.141	0.9779	1min 54s
Seq2Seq Encoder-Decoder model with AM	5.740	0.106	0.9780	1min 59s
Autoencoder	11.609	0.034	0.9044	1min 57s
Transformer (Encoder-Only)	5.904	0.242	0.9782	7min 37s
Transformer (Encoder-Decoder)	11.775	0.321	0.9012	11min 19s

Conclusions and future work

Conclusions and future work

Conclusions:

- We demonstrated that NOTED can reduce duration of large data transfers and improve the efficient use of network resources with production FTS transfers.
- NOTED makes decisions by watching and understanding the behaviour of transfer services → do not need any modification to work with NOTED.
- NOTED may be useful for HL-LHC, if the network bandwidth becomes a limiting factor.
- We currently work with FTS, but if there are other transfer services interested, NOTED could be adapted to them.

Future work:

- Evaluate whether training a single link would perform well on another link → if not, define a training strategy.
- Would it be possible to identify/classify traffic based on historical data and anticipate actions on the network?

Thanks for your attention!

CERN