EPJ Web of Conferences **295**, 07034 (2024) https://doi.org/10.1051/epjconf/202429507034
CHEP 2023

NOTED: An intelligent network controller to improve the throughput of large data transfers in File Transfer Services by handling dynamic circuits

Carmen Misa Moreira[1], *Edoardo* Martelli[1], and *Tony* Cass[1]

[1]CERN - Conseil Européen pour la Recherche Nucléaire, Esplanade des Particules 1, 1211 Meyrin, Geneva, Switzerland, IT department CS group

Abstract. The NOTED (Network Optimised Transfer of Experimental Data) project has successfully demonstrated the ability to dynamically provision network links to increase the effective bandwidth available for FTS-driven transfers between endpoints, such as WLCG sites, by inspecting on-going data transfers and so identifying those that are bandwidth-limited for a long period of time. Recently, the architecture of NOTED has been improved and the software has been packaged for easy distribution.

These improved capabilities and features of NOTED have been tested and demonstrated at various international conferences. For example, during demonstrations at Supercomputing 2022, independent instances of NOTED at CH-CERN (Switzerland) and DE-KIT (Germany) monitored large data transfers generated by the ATLAS experiment between these sites and CA-TRIUMF (Canada). We report here on this and other events, highlighting how NOTED can predict link congestion or a notable increase in the network utilisation over an extended period of time and, where appropriate, automatically reconfigure network topology to introduce an additional or an alternative and better-performing path by using dynamic circuit provisioning systems such as SENSE and AutoGOLE.

1 Introduction

The large scientific data transfers generated by the Large Hadron Collider (LHC) experiments can saturate network links, while alternative paths may be underutilised. Due to the agnostic nature of routing protocols towards network load, we often encounter scenarios where links are congested and yet other expensive links are left idle. Figure 1 shows the congestion suffered on a specific LHC Optical Private Network (LHCOPN) [1] link during a certain period of time.

Network Optimised Transfer of Experimental Data (NOTED) aims to improve network utilisation and better exploit all of the available bandwidth between any given endpoints. The project was introduced in 2020 and presented at several international conferences. First achievements and outcomes are documented in articles [2, 3]. Furthermore, a comprehensive study on traffic forecasting has been conducted using a machine learning approach with Long Short Term Memory (LSTM) neural networks, as detailed in [4].

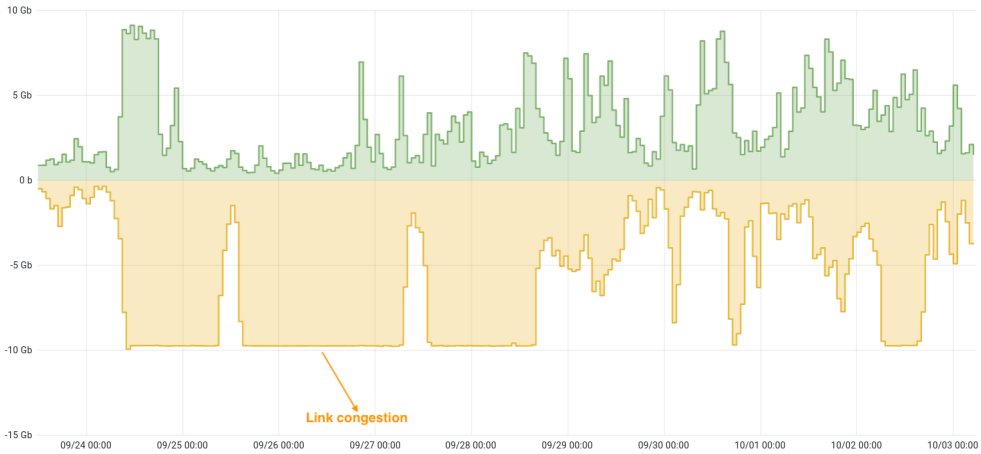


Figure 1: Congestion on an LHCOPN link (incoming - green, outgoing - yellow).

In terms of traffic engineering, the implementation of NOTED presents certain challenges that need to be addressed. Chief among these is that there are frequently situations where a link experiences brief periods of congestion when it is not worthwhile to take any action to redirect traffic to another link in an attempt to alleviate the situation. Accurately estimating the duration of large data transfers is therefore crucial in determining whether it is really beneficial to take actions to exploit underutilised alternative paths and so improve network utilisation and reduce transfer latencies.

Having set out the motivations and objectives of the project, we describe the NOTED architecture and workflow, including the parameters collected from the FTS optimiser and the CRIC repository, in section 2; the packaging of NOTED in section 3; the results of our intensive tests during Supercomputing 2022 (SC22), the International Conference for High-Performance Computing, Networking, Storage and Analysis, in section 4 and, finally, our conclusions and plans for future developments in section 5.

2 Architecture

The File Transfer Service (FTS) [5] and the Computing Resource Information Catalogue (CRIC) [6–8] are the two key services that NOTED uses to retrieve the transfer data and topology information to be used as the basis for making network optimisation decisions. FTS serves as a data transfer service employed by LHC experiments to distribute data to different sites within WLCG (the Worldwide LHC Computing Grid). CRIC is a database used by WLCG sites to declare and expose information about the computing resources accessible at a given site.

NOTED queries FTS at one-minute intervals to retrieve information about on-going and queued transfers. This data is analysed to estimate the duration of transfers and estimate whether an action can be taken to optimise network utilisation, for example by redirecting traffic to alternative links. Given that the network itself is agnostic regarding topology, NOTED relies on the CRIC database to obtain a comprehensive overview of the network elements, encompassing endpoints, sites and federation.

Figure 2 illustrates the NOTED architecture, highlighting the two key components: the transfer broker, serving as the interface that interacts with data transfer applications to retrieve data; and the network intelligence component, which maintains a comprehensive understand-

ing of the network topology and is responsible for undertaking network actions based on the required bandwidth. When inspection of data transfer applications and network monitoring suggests enhanced capacity and bandwidth would be beneficial, this can be achieved by adding alternative links and paths provided by an Software Defined Network (SDN) controller like, for example, SENSE [9–11] for the provision of dynamic circuits or by performing load balancing across existing links.

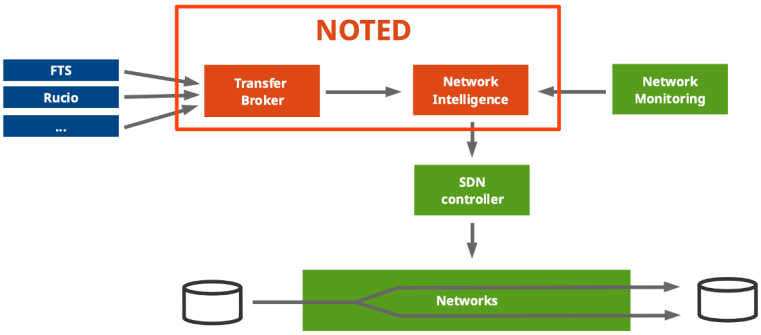


Figure 2: NOTED architecture and components.

The interaction with FTS involves querying the optimiser parameters [12]. The JSON extract below sets out the FTS optimiser parameters employed by NOTED to retrieve data and analyse on-going and queued data transfers. These parameters are crucial for NOTED's network decision-making process, which uses information about the network utilisation and the source and destination of all on-going transfers to calculate the aggregate transfer flow between source and destination pairs. Parameters such as throughput, file size, amount of data, number of parallel transfers, and the submitted transfers that are still waiting in the queue are used by NOTED to compute network utilisation and estimate the duration of the transfers.

JSON sample 1:
FTS optimiser parameters [12].


```
"_index": "monit_prod_fts_raw_queue_state",
"_source": {
    "data": {
        "source_se": "davs://eosatlas.cern.ch",
        "dest_se": "davs://atlas.dcache.nikhef.nl",
        "timestamp": 1685081567365,
        "throughput": 831615682,
        "filesize_avg": 7717212405.757576,
        "success_rate": 100,
        "retry_count": 0,
        "active_count": 5,
        "submitted_count": 0,
        "rationale": "Queue emptying. Hold on.",
        "endpnt": "fts3-pilot.cern.ch",
    },
},
```


Table 1 shows how NOTED aggregates individual on-going transfers into ones that traverse a shared network path. Here there are two aggregate transfers where it can be seen that, despite the throughput in each case increasing from ~50 GB/s to ~80 GB/s, the pending data volume and the number of queued transfers are increasing, a clear sign of congestion.

Table 1: Two on-going bulk data transfers in FTS.

Source endpoint	Destination endpoint	Data [GB]	Throughput [Gb/s]	Parallel transfers	Queued transfers
davs://ccdavatlas.in2p3.fr	davs://webdav.echo.stfc.ac.uk	139.3726	54.0827	453	28557
davs://ccdavatlas.in2p3.fr	davs://webdav.echo.stfc.ac.uk	121.9655	53.6442	422	28538
davs://ccdavatlas.in2p3.fr	davs://webdav.echo.stfc.ac.uk	202.7864	82.0855	862	57880
davs://ccdavatlas.in2p3.fr	davs://webdav.echo.stfc.ac.uk	205.3606	82.0725	888	57790
srm://dcsrm.usatlas.bnl.gov	davs://webdav.lcg.triumf.ca	193.5176	58.8136	530	26294
srm://dcsrm.usatlas.bnl.gov	davs://webdav.lcg.triumf.ca	210.2710	51.0323	567	26314
srm://dcsrm.usatlas.bnl.gov	davs://webdav.lcg.triumf.ca	332.0009	81.7908	905	50152
srm://dcsrm.usatlas.bnl.gov	davs://webdav.lcg.triumf.ca	326.5855	80.1554	903	50028

The workflow of NOTED is divided into three stages. In the first, the network intelligence component queries the CRIC database to gain a comprehensive understanding of the network topology, identifying the relevant endpoints associated with the given source and destination pairs for data transfers. The second stage is where the transfer broker component analyses, every minute, the on-going and queued transfers in FTS and computes the overall network utilisation. Finally, in the third stage, NOTED makes network decisions and, when congestion is detected on a link, either provides a dynamic circuit using an SDN provider such as SENSE to dynamically allocate additional capacity and bandwidth or acts to divert some of the load to existing idle circuits.

YAML sample 2:
NOTED configuration file.

```
src_rcsite: ['src_1', '...', 'src_n']                    ▹ Source
dst_rcsite: ['dst_1', '...', 'dst_n']                    ▹ Destination
events_to_wait_until_notification: x
max_throughput_threshold_link: y                         ▹ [Gb/s]
min_throughput_threshold_link: z                         ▹ [Gb/s]
unidirectional_link: True/False
number_of_dynamic_circuits: k
sense_uuid: 'sense_uuid'                                 ▹ SENSE-O UUID
sense_vlan: 'this is the VLAN description'
from_email_address: 'dedicated email for NOTED'
to_email_address: 'email to send notifications'
subject_email: 'this is the subject'
message_email: "this is the content"
auth_token: auth_token                                   ▹ Authenticantion token
```

The YAML sample above presents an example of the configuration file required to launch NOTED. This allows the network administrator to specify various input parameters, including a list of source and destination sites, threshold values for maximum and minimum throughput triggering congestion detection alarms, essential parameters for the SDN provider, in this case for SENSE, to establish and release dynamic circuits, and additional parameters related to email notification alarms. NOTED will act to optimise network configuration when long-duration transfers exceed the max_throughput_threshold and restore the normal configuration when transfers fall below the min_throughput_threshold; careful selection of these values will prevent repeated reconfigurations in a short time period. The configuration file also enables an administrator to adjust the verbosity level for logging files, choosing between debug, info, or warning messages, as NOTED tracks FTS optimiser events every minute, potentially generating a substantial amount of log data.

3 Package distribution

To facilitate distribution, configuration, installation and deployment, NOTED has been packaged to be compatible with the Ubuntu and CentOS operating systems. Users can access it by downloading the software package from Python Package Index (PyPI), the official software repository for third-party Python applications. Detailed instructions for installation, as well as an example configuration file, are provided alongside the package on PyPI. Installing NOTED from PyPI is as simple as executing the command *pip install noted-dev*. Alternatively, NOTED can be directly obtained from the source [13].

To further simplify distribution and deployment, a Docker image of NOTED based on Alma Linux 9 is provided. Docker enables applications to operate in any environment and operating system, offering isolation from the underlying infrastructure. The installation of NOTED from Docker can be done by executing the command *docker pull carmenmisa/noted-docker*, or alternatively, by directly acquiring it from the source [14].

4 SC22 demonstration and results

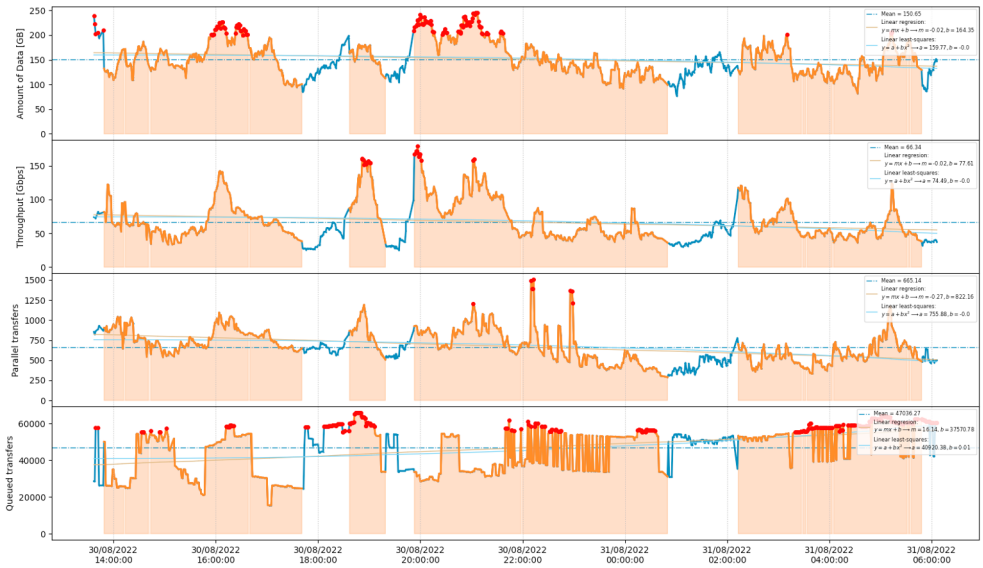
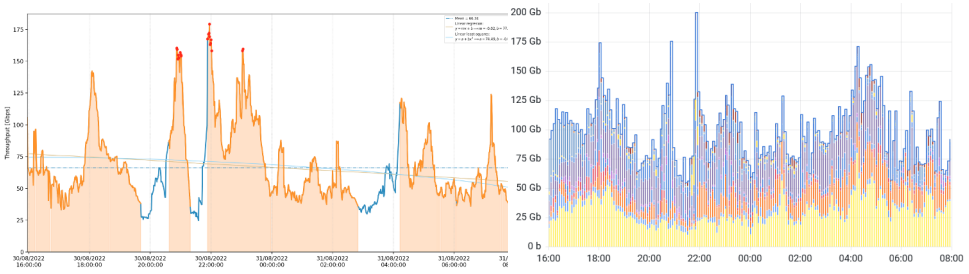


Figure 3: FTS optimiser parameters for WLCG sites in LHCOPN network.

As previously discussed, NOTED retrieves parameters from the FTS optimiser at one-minute intervals to monitor and track data transfers and raises an alarm when it detects potential long-term congestion in a link. Figure 3 shows the main parameters used by NOTED to determine if there is congestion and estimate the duration of transfers. These parameters are: the amount of data (1st row), throughput (2nd row), number of parallel transfers (3rd row) and queued transfers (4th row); the figure shows the data collected from FTS between the 30th and 31st of August 2022. During these days, all WLCG sites within the LHCOPN network were being monitored, with a maximum threshold of 80 Gb/s and a minimum threshold of 40 Gb/s for providing and releasing dynamic circuits. The blue line of the graph corresponds to events and parameters generated by the FTS optimizer, collected every minute, and the orange colour indicates the time window during which the dynamic circuit was provided. It can clearly be seen that NOTED identified long-term link congestion, and estimating that an action could usefully be taken in the network to increase the available bandwidth by providing a dynamic circuit, successfully improved overall transfer performance for the end users.

For verification and validation of the work carried out, the results and observations of NOTED concerning the network utilisation were compared with those reported by the CERN LHCOPN production routers. Figure 4 demonstrates that the throughput reported by the FTS optimiser, which NOTED uses as a parameter to detect link congestion and generate alarms, corresponds with the reported values by the CERN LHCOPN production routers. Therefore, we can conclude that by inspecting on-going data transfers in FTS, it is possible to get an understanding of network usage, enhance its performance and optimise resource utilisation by executing actions on the network topology, to add more bandwidth and capacity.

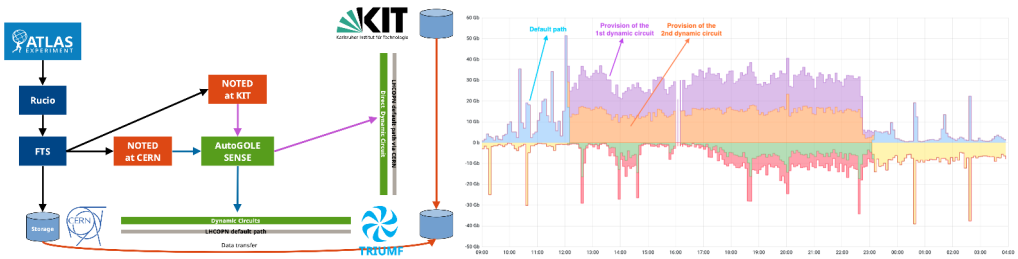


(a) Throughput reported by the FTS optimizer that NOTED used to generate the alarms.

(b) Throughput reported by the LHCOPN production border router.

Figure 4: Throughput used by WLCG sites in LHCOPN network.

From 13th to 15th of November 2022, a wide range of intensive tests of NOTED were conducted at the International Conference High-Performance Computing, Networking, Storage and Analysis Conference, also known as SC22. Figure 5 (a) shows a diagram of the demonstration carried out during SC22, highlighting the participants and collaborators involved. As can be seen, CH-CERN and DE-KIT served as the entities where an instance of NOTED ran as a network controller and also provided data storage. CA-TRIUMF provided data storage on the opposite side of the link, with ESnet [15] enabling the provision of dynamic circuits between [CH-CERN, CA-TRIUMF] and [DE-KIT, CA-TRIUMF] in collaboration with CANARIE [16], STARLIGHT [17] and SURF [18]. NOTED was configured to raise an alarm and provide dynamic circuits when the throughput exceeded 40 Gb/s and release them when the throughput dropped below 20 Gb/s. For this particular demonstration, the ATLAS experiment was responsible for generating the large data transfers between the participating sites. At this point, it can be mentioned that we configured two instances of NOTED for reasons of distribution to different institutions, but it would be sufficient to use a single instance.



(a) Diagram of the SC22 demonstration.

(b) Results of SC22 demonstration.

Figure 5: NOTED demonstration conducted at SC22.

The results of the NOTED demonstration at SC22 are presented in figure 5 (b). The blue and yellow graphs represent the traffic routed through the original link, while the purple and orange graphs show the two dynamic circuits that were provisioned during the demonstration. The graph shows that around 12:00 AM on the 15th of November 2022, NOTED detected a high demand for network capacity and estimated that the duration of the transfer would be for a long period of time. Consequently, two dynamic circuits were provided and the traffic was redirected through them. Subsequently, at around 10:30 PM on the same day, NOTED released the circuits as network usage returned to normal and there were no more data transfers to cause congestion in the link. Through the utilisation of these dynamic circuits, traffic load balancing was achieved, and the network topology was dynamically modified by providing alternative links that were not present in the original network configuration.

5 Conclusions

In conclusion, we can say that NOTED can be used to reduce the duration of large transfers in conditions where the available bandwidth is not sufficient and is causing congestion by adding additional capacity from dynamic circuit provisioning services or by redirecting part of the traffic to an existing idle link.

The major need now is to improve the accuracy of traffic forecasting, especially for the duration of larger data transfers, and so improve the effectiveness of the decision-making algorithm in terms of its accuracy in identifying when to request and drop dynamic circuits.

References

[1] E. Martelli, S. Stancu, *Lhcopn and lhcone: Status and future evolution* (2015), https://dx.doi.org/10.1088/1742-6596/664/5/052025

[2] C. Busse-Grawitz, E. Martelli, M. Lassnig, A. Manzi, O. Keeble, T. Cass, *The noted software tool-set improves efficient network utilization for rucio data transfers via fts* (2020), https://api.semanticscholar.org/CorpusID:229255607

[3] J. Waczynska, E. Martelli, E. Karavakis, T. Cass, *Noted: a framework to optimise network traffic via the analysis of data from file transfer services* (2021), https://doi.org/10.1051/epjconf/202125102049

[4] J. Waczynska, E. Martelli, S. Vallecorsa, E. Karavakis, T. Cass, *Convolutional lstm models to estimate network traffic* (2021)

[5] Karavakis, Edward, Manzi, Andrea, Arsuaga Rios, Maria, Keeble, Oliver, Garcia Cabot, Carles, Simon, Michal, Patrascoiu, Mihai, Angelogiannopoulos, Aris, *Fts improvements for lhc run-3 and beyond* (2020), https://doi.org/10.1051/epjconf/202024504016

[6] Anisenkov, Alexey, Andreeva, Julia, Di Girolamo, Alessandro, Paparrigopoulos, Panos, Vasilev, Boris, *Cric: Computing resource information catalogue as a unified topology system for a large scale, heterogeneous and dynamic computing infrastructure* (2020), https://doi.org/10.1051/epjconf/202024503032

[7] Anisenkov, Alexey, Andreeva, Julia, Di Girolamo, Alessandro, Paparrigopoulos, Panos, Vedaee, Aresh, *Cric: a unified information system for wlcg and beyond* (2019), https://doi.org/10.1051/epjconf/201921403003

[8] M. Alandes, J. Andreeva, A. Anisenkov, G. Bagliesi, S. Belforte, S. Campana, M. Dimou, J. Flix, A. Forti, A. di Girolamo et al., *Consolidating wlcg topology and configuration in the computing resource information catalogue* (2017), https://dx.doi.org/10.1088/1742-6596/898/9/092042

[9] I. Monga, C. Guok, J. MacAuley, A. Sim, H. Newman, J. Balcas, P. DeMar, L. Winkler, T. Lehman, X. Yang, *Software-defined network for end-to-end networked science at the exascale* (2020), https://www.sciencedirect.com/science/article/pii/S0167739X19305618

[10] J. Guiang, A. Arora, D. Davila, J. Graham, D. Mishin, I. Sfiligoi, F. Wuerthwein, T. Lehman, X. Yang, C. Guok et al., *Integrating end-to-end exascale sdn into the lhc data distribution cyberinfrastructure* (2022), https://doi.org/10.1145/3491418.3535134

[11] T. Lehman, X. Yang, C. Guok, F. Wuerthwein, I. Sfiligoi, J. Graham, A. Arora, D. Mishin, D. Davila, J. Guiang et al., *Data transfer and network services management for domain science workflows* (2022), 2203.08280

[12] *Fts optimiser documentation*, https://fts3-docs.web.cern.ch/fts3-docs/docs/optimizer/optimizer.html

[13] C. Misa, E. Martelli, *Noted pypi source code*, https://pypi.org/project/noted-dev/

[14] C. Misa, E. Martelli, *Noted docker source code*, https://hub.docker.com/r/carmenmisa/noted-docker

[15] *Energy science network, lawrence berkeley national laboratory, united states department of energy (doe), united states national research and education network (nren)*, https://www.es.net/

[16] *Canada national research and education network (nren)*, https://www.canarie.ca/

[17] *Starlight: the optical star tap*, https://www.startap.net/starlight/

[18] *Dutch national research and education network (nren)*, https://www.surf.nl/en