International Journal of Modern Physics A Vol. 36, No. 5 (2021) 2130005 (13 pages) © World Scientific Publishing Company DOI: 10.1142/S0217751X21300052

TRACER (TRACe route ExploRer): A tool to explore OSG/WLCG network route topologies

Evgeniy Tretyakov^{*} and Alexey Artamonov[†]

National Research Nuclear University MEPhI, 31 Kashirskoe shosse, Moscow, Russia Plekhanov Russian University of Economics, 36 Stremyanny Lane, Moscow, Russia * estretyakov@mephi.ru † AAArtamonov@mephi.ru

Maria Grigorieva[‡]

Lomonosov Moscow State University, Leninskie Gory 1, bld. 4, Moscow, Russia Plekhanov Russian University of Economics, 36 Stremyanny Lane, Moscow, Russia maria.grigorieva@cern.ch

Alexei Klimentov

Brookhaven National Laboratory, Upton, New York, USA aak@bnl.gov

Shawn McKee

University of Michigan, 450 Church St, Ann Arbor, Michigan, USA smckee@umich.edu

Ilija Vukotic

University of Chicago, Chicago, Illinois, USA ivukotic@uchicago.edu

> Received 28 July 2020 Revised 17 November 2020 Accepted 21 December 2020 Published 10 March 2021

The experiments at the Large Hadron Collider (LHC) rely upon a complex distributed computing infrastructure (WLCG) consisting of hundreds of individual sites worldwide at universities and national laboratories, providing about half a billion computing job slots and an exabyte of storage interconnected through high speed networks. Wide Area

[‡]Corresponding author.



E. Tretyakov et al.

Networking (WAN) is one of the three pillars (together with computational resources and storage) of LHC computing. More than 5 PB/day are transferred between WLCG sites. Monitoring is one of the crucial components of WAN and experiments operations. In the past years all experiments have invested significant effort to improve monitoring and integrate networking information with data management and workload management systems. All WLCG sites are equipped with perfSONAR servers to collect a wide range of network metrics. We will present the latest development to provide the 3D force directed graph visualization for data collected by perfSONAR. The visualization package allows site admins, network engineers, scientists and network researchers to better understand the topology of our Research and Education networks and it provides the ability to identify nonreliable or/and nonoptimal network paths, such as those with routing loops or rapidly changing routes.

Keywords: Networks; visualization; perfSONAR; traceroute; topology.

PACS number: 07.05.Bx

1. Introduction

Computer networks are extremely important for data-intensive science, enabling the construction of complex distributed computing infrastructures that allow scientists to access data and computing resources locally, regionally, nationally and internationally. Network technologies enable collaborative research and data mobility and access around the world. Because of the foundational role that networks play in modern science, it is critical we ensure their high reliability and stability, but this is often very challenging. Networks by their nature cross multiple administrative domains, meaning that there is no single owner or manager of a given end-to-end network path. When soft failures (degradation of performance but not loss of connectivity) occur somewhere in the network, it can be very difficult to identify and localize the problem, leading to periods of lost scientific productivity across distributed resources which can last weeks to months and even to years, in some cases.

The particular network infrastructure that we will focus on in this paper is the one used by the scientific experiments at the LHC. The Worldwide LHC Computing Grid (WLCG) project¹ — a global collaboration of around 170 computing centers in more than 40 countries, is constructed on top of a conglomeration of local, regional, national and international networks, primarily research and education (R&E) networks but including some commercial segments. The mission of the WLCG project is to provide global computing resources to store, distribute and analyze peta- and exabyte-scale LHC data for the LHC experiments.

The LHC experiments at CERN currently transfer a few petabytes of data daily, in aggregate. While each data transfer is characterized by a single source and destination, the network path that is actually traversed between those endpoints can vary, based upon routing protocols and the status of the network components. The critical point to note here is that when there is a network problem, the first thing we need to know is "what path did the problem occur on?" Answering these questions identifies the relevant elements that could be the location for the problem being observed. Fortunately, tracking of these dynamically changing paths or routes has become possible thanks to the massive and detailed network measurements being provided for the last several years by the OSG-LHC/WLCG network monitoring activity.² While the creation and operation of this network metric pipeline has given us the data we could use to find and localize network problems, in practice, it has given us an additional challenge in how best to identify, filter and visualize the massive amount of path information we are monitoring. This is the motivation for the creation of TRACER (TRACe route ExploRer)³ and in the rest of this paper we will outline the technical details of TRACER and demonstrate how it is being used with OSG/WLCG data.

2. Background

In this section, we describe some of the relevant concepts and technologies used to define and measure network paths. The IP network path between a source and destination is determined by devices called routers, which operate at the IP layer (layer 3) in the OSI network stack. A network path can be described by the sequence of router IP addresses that packets traverse as they move from the source to the destination. Each router determines which next router to forward a network packet to (usually based only upon the destination network) using a routing table. Routing tables may be constructed automatically or manually. Manual routing tables comprises static routes while automatic produces dynamic routes. The routing table consists of at least three elements:

- (1) network ID the destination network;
- (2) metric a routing metric indicating the route priority;
- (3) gateway the address of next router to send the packet to.

WLCG routing tables are maintained semi-automatically, in most cases, using suitable software tools and frameworks. Even so, providing high performance at peak efficiency and availability becomes challenging for network operation staff. Even the slightest amount of packet loss on large bandwidth-delay paths drastically reduces the throughput achievable.⁴ We note that it is possible to improve the behavior of TCP in the presence of packet loss but using different congestion control algorithms like TCP-BBR.⁵ Maintaining networks in distributed computing systems would be impossible without monitoring and performance analysis tools. Network operation staff depend upon such tools to identify and fix network issues.

Toolkits, such as perfSONAR⁶ (performance Service-Oriented Network monitoring ARchitecture), are used in WLCG for network testing and monitoring.⁷ WLCG has deployed perfSONAR globally across its sites and has configured it to automatically monitor paths of interest, collecting and archiving network performance metrics in an efficient way. In 2020, we have more than 250 perfSONAR instances installed on 130 sites, constantly measuring and retrieving topology information, network delay, packet loss and bandwidth of network routes. The data retrieved by perfSONAR tools are studied and analyzed in order to troubleshoot network performance problems. Machine learning methods have been applied for the prediction of performance issues based upon packet loss data measured by perfSONAR.⁸ To better understand the data and identify issues in our networks, we rely on dashboards and visualization tools like Kibana and Grafana, both standard for data stored in Elasticsearch.⁹ Kibana enables interested users to visualize data with histograms, line graphs, pie charts, sunbursts, tag clouds and more. Grafana is a powerful tool to query, visualize, and gain insights from data. Both of these tools are important components of WLCG network monitoring.¹⁰

Even with the monitoring, analysis and visualization tools, we have already described, we faced a challenge in analyzing and visualizing our measured network topologies. The problem is the complexity and dynamic nature of the path data. The data describing the network paths comes from either traceroute or tracepath tools and consists of the list of IP addresses of the routers along the path from source to destination, as well as additional information about the round-trip-time (rtt) to each router, the Autonomous System (AS) number which designates the network entity which manages each router, and, for tracepath, the maximum MTU supported to each router. Each router along the path constitutes a hop and the number of hops, even between a specific source and destination, can vary as conditions in the network change. Because our existing tools (Kibana, Grafana and Elasticsearch) were not well matched to the challenge of visualizing and understanding the network path data, we needed to develop TRACER. This paper describes TRACER, a 3D forcedirected graph visualization for network routes. TRACER includes a web interface with filtering, data preparation functions and interactive visualization that allows users to monitor and explore network paths and how they change in time.

As TRACER was being developed, we engaged with the perfSONAR developer team, presenting early prototypes to get feedback on. The perfSONAR developers were enthusiastic about the new visualization capability for traceroute data and have had something similar on their future development list for many years. Given that the next release of perfSONAR (v4.4) will focus on visualization, we may have an opportunity to collaborate with them on improving integration between TRACE and perfSONAR.

3. Visualization Pipeline

Typical applications for visual analysis leverage specific visualization pipelines that include four main stages: sourcing, filtering, mapping and rendering. The sourcing stage provides the access to initial data sources. The filtering stage allows a user to select specific data for further analysis. The mapping stage is where filtered data items are mapped to geometric primitives (i.e. points, lines, spheres, cylinders) and their attributes, such as color, position or size. Mapping is the most critical stage for the efficiency and quality of the resulting visualization. The rendering stage is the process of transformating the geometric data to graphical images. route visual participation and the sections of the participation of the provides well-structured easy to maintain source code. The appendix of our visualization pipeline physical action architecture is presented in Figure 1. Figure 2 shows the main window of

the TRACER web application. It consists of central frame for route visualization,



Fig. 1. The 3-D force directed graph visualization application architecture Fig. 1. The 3D force directed graph visualization application architecture.



Fig. 2. TRACER Web interface

The 3D force-directed graph visualization¹¹ includes both client and server components. For TRACER the server is based on Django which acts as an application programming interface (API) for the Elasticsearch data. The server accepts search queries and returns data retrieved from the Elasticsearch. The client provides a graphical user interface (GUI) to generate search queries and analyze and display the resulting data. The complex GUIs are built using the React JavaScript framework, which provides well-structured, easy to maintain source code. The application architecture is presented in Fig. 1. Figure 2 shows the main window of the TRACER web application. It consists of central frame for route visualization, a left frame showing the filtering options, a table of filtered paths below the main route visualization, and a right pane showing path details. In the following sections, we detail the four main stages of our visualization pipeline.

3.1. Sourcing

The data from WLCG perfSONAR servers is collected and transported to the Elasticsearch analytics cluster^a at the University of Chicago. A sample of a network path record is shown in Figure 3.



The data record schema for the Network Route Visualization Tool. Fig. 3.

Fig. 3. The data record schema for the Network Route Visualization Tool.

3.1. Sourcing

The addra changed and Ale and a second a s Elesticseanshraicallaticsciluster tintehawinepsickets Chicago elweemple sofa active have the destinational the pass through a number of intermediate network devices (routers). Eackatha ecord ckergidses not work more worked a source for the station of the second Path, atequipational received inclinants When an ackats of a value beta points sand a living as klestizetienethen pesutersuch zepundere ofdertenned digt (potyppk cleadere seeut Dadh Easehreizerte sozielert no asservized Constantion and the set of t sThus network at record contains metadata about the endpoints and all hops between them. All routers in a path are identified by IPv4/IPv6 addresses. Each data General stored matip WLCG Elasticsearch is composed of the following three sets of parameters Measurement Archiver (MA) that provided the original data.

- <u>*imestamp*</u> the time of the perfSONAR measurement.
 - n_hops the number of hops on a route.
 - $\underline{MA_{max}}$ The Measurement Archiver that provided the original data.
 - timestamn P protocol type (true if HV6, false if HV4).

 - \underline{n} -hops \underline{m} the number of hops on a route \underline{n} dentifying a network path. $max_{rtt} - maximum$ value of the round trip time (rtt).
- ^ahttps://atlas-kibana.mwt2.org.5601 (true if IPv6, false if IPv4).

route_sha1 — cryptographic hash uniquely identifying a network path.

• Endpoints Parameters

src_site, *dest_site* — source and destination WLCG site names.

^ahttps://atlas-kibana.mwt2.org:5601/

- *src_host*, *dest_host* source and destination host names.
- -ipv6 version of IP protocol (ipv6 = True/False).
- *src_VO*, *dest_VO* virtual organization of the src/dest WLCG owner.^b
- *src_production, dest_production* Boolean indicating source and destination production status (test versus production).

• Router Parameters

- round-trip time (rtts) the vector of round-trip-times (rtts) in ms to reach each router along the path.
- time to live (ttls) the vector of the values of the time-to-live field, decremented as a packet transits each router.
- autonomous system number (asns) showing who operates each router.

The TRACER tool allows us to explore this network route visually. Each route measurement is stored in the Elasticsearch as a single document. There are couple of parameters that require further clarification:

- To identify unique paths through the network, we hash the vector of hops using the SHA1 hashing function to create route-sha1. This identifier allows us to easily track or aggregate on that specific path when measured by traceroute.
- The src_production or dest_production Booleans indicate whether a site is in production (true) or being tested (false) and is not related to the measurement.

3.2. Filtering

The filtering functions allow construction of search queries on the client-side. The filtering enables users to select time ranges, IP version (IPv4 or IPv6) and specific endpoint values: Sources (*src*), Source hosts (*src_host*), Destinations (*dest*), Destination hosts (*dest_host*). The filtering widgets are placed on the left side of the application window. The user query construction results in a query in JSON format compatible with the Elasticsearch engine. This query is passed from client to server where the Django application gets records from Elasticsearch and prepares the data for further graph building on the client.

3.3. Mapping

Five types of nodes are used to build force directed graph visualization. These node types are

- Source domain name *src_host*,
- Source IP address *src*,
- Router IP addresses *hops*,
- Destination IP address *dest*,
- Destination domain name dest_host.

^bhttps://wlcg.web.cern.ch/virtual-organisations

ce and destination at the same time, then the node's color is set to put es of *dest_host* and *dest* can also be highlighted in red when a path does h the actual destination (an incomplete path).

⁸_{E.} Tretyakov, et al. Tretyakov et al.

≸R6

50 58:20.192

6

.192

148.187.129.15

marperf02.in2...

58.20.192 148.187.129.15 marperf02.in2... perfsonar01.l.

perfsonar02.l.

<

58.20 192

by PRINCETON UNIVERSETY



Fig. 5. Table with filtered paths.

2 /2

Showing 6 to 9 of 9 (9) records

IN2P3-CPPM

ATLAS

ATLAS

>

ATLAS

ATLAS

 \odot

 \odot

з



fluctuating across many distinct paths. One cause of the fluctuation is the use of Equal-cost multi-path routing $(\text{ECMP})^{13}$ which some networks use to distribute load across multiple paths to a destination.

Figure 9 presents a complete network path between two perfSONAR servers.





FigFig. 9NeNetwork and the between psl. hpc utfism chard Φ 90000245. 45 the psite sites fluctuating across many distinct paths. One cause of the fluctuation is the use of Equal-cost multi-path routing (ECMP)¹³ which some networks use to distribute

4. Reventions and the spaces have deissing dops. One interesting feature to note in this path is the showed by the space of the space o



Figure 9 presents a complete network path between two perisonals servers. This path Can⁹beNconsidered a stable because dis tas moth been ichanged for quite some time and does not have missing hops. One interesting feature to note in this path 98 think the from 95 tween 2000 1204. For (Node 3) fait 200 09 204 53 (Node 4) routers is 113 to milliseconds. These two routers are both part of the RedCLARA routers is 113.60 milliseconds. These two routers are both part of the RedCLARA network, with the first in Chile and the second in Mexico. The second large RTTS is due to a phonoaccess the Atlantic Central

Figureulo illustrateesthe path with self-poped segment indink houdpat happen duridging name downing. We analyzed all aparts downer in conter (Buter) (P) appears more clighthones in the path and if if is found we work high the adopt is 150 ping. The Fig. 10. Network path between CA-SFU-T2 and INFN-MILANO-ATLASC sites.



Fig. 10. Network path between CA-SFU-T2 and INFN-MILANO-ATLASC sites. Fig. 10. Network path between CA-SFU-T2 and INFN-MILANO-ATLASC sites.

2130005-10

router with IP 205.189.32.54 is encountered twice on the route. TRACER allows us

1

by routing misconfigurations or when internet routes are in the process of changing or updating.

Figure 11 illustrates case when certain routers, due to their configuration or a technical malfunction, do not respond to a request from perfSONAR that measures the network path. Loss of network path data is observed form 9th hop to the desti-TRACER (TRACE route ExploRer): A tool to explore OSG/WLCG



Fig. 11. Network path between RAL-LCG2 and NCP-LCG2 with missed nodes. Fig. 11. Network path between RAL-LCG2 and NCP-LCG2 with missed nodes.

router with IP 205.189.32.54 is encountered twice on the route, TRACER allows us nation, because relevant routers do.not, respond with repeated measurements. But to spot such self-looped segments visually easily. Such looping paths can be caused the path itself reaches the destination, thus it can be assumed that these losses by routing misconfigurations or when internet routes are in the process of changing occur due to the security settings of the respective routers.

Figure 11 illustrates the case when certain routers, due to their configuration or a **5. Future Work** technical malfunction, do not respond to a request from perfSONAR that measures This first-ordepath ILBsAGERepworkdepathfulactions berief from perfSONAR that measures this first-ordepath ILBsAGERepworkdepathfulactions berief from perfSONAR that measures standscip at inits of the source of the standard state of the second depath of the second depa

blax dynamic topologies measured by the global deployment of perfSONAR inTRACER can be very slow or will not even respond if the user selects a time stances in hundreds of locations worldwide. Its primary use-case is to support users, . period of more than a few minutes without selecting additional filtering.
TRACER currently displays only data about network traces, but a very interwork topology involved in our WLCG minastructure. However, as with almost any esting extension, would be to allow it, to gather and display associated metrics along a set of identified naths, e.g., latency, throughput and packet-loss.
TRACER could be used to explore how paths change in time, and updating to work on in the future: the user interface to support this would be a valuable extension.

- TRACER can be very slow or will not even respond if the user selects a time period of more than a few minutes without selecting additional filtering.
- TRACER currently displays only data about network traces, but a very interesting extension would be to allow it to gather and display associated metrics along a set of identified paths, e.g. latency, throughput and packet-loss.
- TRACER could be used to explore how paths change in time, and updating the user interface to support this would be a valuable extension.

E. Tretyakov et al.

- Visualization of the intersections of traces of multiple source-destination pairs where some issues were detected allows us to visually understand path correlations. Additionally, TRACER may be used for the visualization of all network paths that use one particular router.
- Identifying asymmetric routes which can be associated with network configuration problems. The challenge here is in two parts: (1) we need to identify all the IP addresses routers utilize, since a given router may report a different IP depending upon how it is traversed and (2) we need to improve the visualization to better support this use case (similar to the previous item above) so we can clearly visualize source-to-destination and destination-to-source to identify asymmetries.
- TRACER visualizes routes in terms of router hops but we also have data about the AS number (network owner) and this can allow us to provide a much more condensed view of the path, highlighting the major network entities along the path.

With the updated, refactored code in this first release, we believe that much of the above should be quickly achievable and expect at least few of the items above will be addressed in our next release.

6. Conclusion

We have described the motivation, design and implementation of TRACER, a tool for visual exploration of network topology measured by the global deployment of OSG and WLCG perfSONAR toolkit instances. TRACER is now integrated in CERN computing infrastructure and can be used for the near real-time network monitoring and retrospective analysis of network faults and malfunctions. An additional use-case is interactive visualization to support the development of mathematical and statistical methods for the detection of network failures.

Acknowledgments

The development of the visualization platform is financially supported by the Russian Science Foundation award #18-71-10003 (research conducted in Lomonosov Moscow State University). We also gratefully acknowledge the Russian Science Foundation award #19-71-30008 (for research related to the Russian scientific data lake prototype, this research is conducted in Plekhanov University, Moscow) and the US National Science Foundation which supported data collecting, storing and processing part of this work through NSF grants #1836650 and #1827116.

References

 Y. Bird, P. Buncic, F. Carminati, M. Cattaneo, P. Clarke, I. Fisk, M. Girone, J. Harvey, B. Kersevan, P. Mato, R. Mount and B. Panzer-Steindel, Update of the Computing Models of the WLCG and the LHC Experiments, Tech. Rep. CERN-LHCC-2014-014.LCG-TDR-002, CERN (Apr 2014).

- M. Babik, S. McKee, B. P. Bockelman, E. M. F. Hernandez, E. Martelli, I. Vukotic, D. Weitzel and M. Zvada, Improving WLCG networks through monitoring and analytics, in *EPJ Web Conf.*, Sofia, Bulgaria (2019), p. 08006.
- E. Tretyakov and I. Vukotic, sand-ci/trace-explorer-tracer: First release, doi:10.5281/ zenodo.3903716 (2020).
- 4. T. V. Lakshman and U. Madhow, IEEE/ACM Trans. Network. 5, 336 (1997).
- M. Hock, R. Bless and M. Zitterbart, Experimental evaluation of bbr congestion control, in 2017 IEEE 25th Int. Conf. on Network Protocols (ICNP) (2017), pp. 1–10.
- B. Tierney, J. Boote, A. B. E. Boyd, M. Grigoriev, J. Metzger, M. Swany, M. Zekauskas and J. Zurawski, Instantiating a global network measurement framework, in SOSP Workshop on Real Overlays and Distributed Systems (ROADS'09), Big Sky, MT, USA, 2009.
- S. Campana et al., J. Phys.: Conf. Ser. 513, 062008 (2014), doi:10.1088/1742-6596/ 513/6/062008.
- 8. M. Babik and H. Borras, Simulating network throughput by correlating perfSONAR measurements with link utilisation, Tech. Rep. CERN-IT-Note-2017-001, CERN (2017).
- 9. Elastic-Developers, Elastic stack, https://www.elastic.co/guide/index.html, (2018).
- S. McKee, Prototype OSG/WLCG kibana network dashboards, https://atlaskibana.mwt2.org/s/networking/app/kibana#/dashboard/07a03a80-beda-11e9-96c8d543436ab024?_g=() (2020).
- 11. Wikipedia contributors, Force-directed_graph_drawing Wikipedia, the free encyclopedia, https://en.wikipedia.org/wiki/Force-directed_graph_drawing (2020).
- 12. Wikipedia contributors, Hop (networking) Wikipedia, the free encyclopedia, https://en.wikipedia.org/wiki/Hop_(networking) (2020).
- 13. Wikipedia contributors, Hop (networking) Wikipedia, the free encyclopedia, https://en.wikipedia.org/wiki/Equal-cost_multi-path_routing (2020).