Identifying and Understanding Scientific Network Flows

Garhan Attebury[8], *Marian* Babik[2], *Dale* Carder[3], *Tim* Chown[6], *Andrew* Hanushevsky[4], *Bruno* Hoeft[7], *Andrew* Lake[3], *Michael* Lambert[9], *James* Letts[8], *Shawn* McKee[1], *Karl* Newell[10], and *Tristan* Sullivan[5]
for the Research Networking Technical Working Group,*

[1]University of Michigan Physics, 450 Church St, Ann Arbor MI 48109, USA
[2]European Organisation for Nuclear Research (CERN), Geneva, Switzerland
[3]Lawrence Berkeley National Laboratory, Berkeley, CA ,USA
[4]SLAC National Accelerator Laboratory, Menlo Park, CA, USA
[5]University of Victoria, Victoria, British Columbia, Canada
[6]Jisc, Bristol, UK
[7]Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany
[8]University of California, San Diego, CA, USA
[9]University of Nebraska, Lincoln, NE, USA
[10]Pittsburgh Supercomputing Center, Carnegie Mellon University, Pittsburgh, PA, USA
[11]Internet2, Ann Arbor, MI, USA

Abstract. The High-Energy Physics (HEP) and Worldwide LHC Computing Grid (WLCG) communities have faced significant challenges in understanding their global network flows across the world's research and education (R&E) networks. This article describes the status of the work carried out to tackle this challenge by the Research Technical Networking Working Group (RNTWG) and the Scientific Network Tags (Scitags) initiative, including the evolving framework and tools, as well as our plans to improve network visibility before the next WLCG Network Data Challenge in early 2024. The Scitags initiative is a long-term effort to improve the visibility and management of network traffic for data-intensive sciences. The efforts of the RNTWG and Scitags initiatives have created a set of tools, standards, and proof-of-concept demonstrators that show the feasibility of identifying the owner (community) and purpose (activity) of network traffic anywhere in the network.

1 Introduction

High-Energy Physics (HEP) experiments rely on networks as one of the critical components of their infrastructure. These networks are used to interconnect participating sites, data centres, and scientific instruments within laboratories and throughout the world. The network traffic used by the HEP and WLCG experiments is split across purpose-built networks such as LHCOPN[1] and LHCONE[1] as well as the research and education networks (R&E networks) operated by campuses, research organisations and national research and education networks (NRENs) and general public networks. Managing such global networks, which support data-intensive sciences, is becoming increasingly complex because of the increasing

[*] RNTWG Charter available at https://cern.ch/rntwg-charter.

number of science projects simultaneously using the same infrastructure resources, the increase in the total amount of data transferred, and the divergent variables that govern these projects, such as requirements, constraints, configurations, and technologies. One of the key challenges in tackling this complexity is improving the current insufficient visibility into which scientific communities are creating the network flows and the purpose of these flows. Without this visibility, it is difficult to understand how the network is being used, what is the actual community usage across different network segments, how to optimise network performance, and how to debug and troubleshoot issues that arise during operations.

The Research Networking Technical Working Group (RNTWG) was formed in the spring of 2020, partially in response to this challenge. The first of its three working areas concerns network visibility; specifically, exploring the use of new packet marking and flow labelling techniques (as defined below) to identify the owner and associated activity of the network traffic. The resulting Scitags initiative (https://www.scitags.org/) was created to develop the generic framework and standards and to push the proposed model into production, not just for HEP/WLCG, but for any global user of the R&E networks.

This paper describes the status of the work to date, including the evolving framework and tools, as well as our plans to get this capability into production.

2 Framework

The main goal of the Scitags initiative is to develop a framework that would help identify communities and their activities at the network level. In the WLCG case, this would mean identifying the experiments (ATLAS, CMS, etc.) and their high-level activities (such as production, analysis, data challenge, etc.) so that network providers can collect this information and correlate it with the other data they have available.

Network traffic typically takes the form of a set of network flows, which is a sequence of packets that, without exceeding a given threshold time, share the same source and destination addresses, as well as the same transport protocol and port numbers. A file transfer typically uses one or more network flows, often in parallel.

To be able to identify traffic as belonging to a particular combination of community and activity, we use a **flow identifier**, which in the WLCG case is a tuple of experiment and activity. The community and activity values to be used are maintained in the Scitags registry.

Two new Scitags mechanisms are proposed to support the identification of the community and activity associated with network traffic:

- *Packet marking* is the process of adding a tag to every network packet sent, identifying the community and activity to which it belongs. In this case, the flow identifier is added to the header of each packet.
- *Flow labelling* uses a separate communication channel, which carries both the flow identifier and the meta-data to identify the flow of the reference network.

After conducting initial feasibility studies, Scitags initiative has proposed a framework that can provide high-fidelity visibility for data-intensive scientific communities. The WLCG project is used as a real-world example to demonstrate how such functionality can be implemented and deployed in production environments. The proposed framework is based on the following rationale:

- The framework is open and can be used by any data-intensive science community, both HEP and non-HEP. Since network resources are inherently shared (even for dedicated L3VPNs such as LHCONE) and, with new scientific experiments coming online in the near future, it is important to have a framework that is widely applicable and accessible to different communities.

- Identify the owner and purpose of the traffic by using either Scitags packet marking or flow labelling.
- Decouple producers (storages and/or compute infrastructure) and consumers (collectors and receivers) by defining a standard for the exchange of flow identifiers. This allows for greater flexibility and scalability, as producers and consumers can independently evolve and scale.
- Use coarse definitions of community/activity to provide insight into the aggregate. The aim is to capture most of the overall traffic from the community.
- Enable tracking and correlation with existing network flow monitoring (IPFIX, sFlow, etc.). This enables the use of existing tools and infrastructure, reducing the cost and complexity of implementation for network providers.
- Quantify global behaviour and analyse trade-offs at scale, e.g. data-set & storage placement, job scheduling, etc.
- The framework also has the potential, particularly with Scitags packet marking, to be used for traffic engineering in the future. This would allow for the optimisation and potential separation of the network traffic for each community.

In the Scitags framework, *flow labelling* is implemented using a separate communication channel in which UDP packets (so-called UDP fireflies) are sent alongside the real network traffic. UDP fireflies carry syslog-formatted information containing the community and its activity, as well as the information required to identify the reference network traffic. Flow labelling works for both IPv4 and IPv6, and is easier to implement as simple socket access is sufficient, but requires more complex collection and correlation mechanisms (dedicated collectors for specific packets, port mirroring, etc.).

Packet marking is implemented by encoding the community and its activity within the 20-bit Flow Label field in the IPv6 packet header and is thus an IPv6-specific solution. While packet marking is a more direct way to identify traffic, it is more complex to implement (since it requires interaction with the Linux kernel), but since each packet is a carrier of the flow identifier, it is easier to extract, as network equipment typically offers such capability. It is possible that other packet marking approaches will also be evaluated, e.g. using the IPv6 Hop-by-Hop option.

The core elements of the framework and their interaction are shown in Figure 1 and work as follows:

- Data transfer agent: any entity that generates network traffic and implements either Scitag flow labelling or packet marking (or both). In WLCG, this would typically be storage systems (EOS, XRootD, dCache, StoRM, Echo, etc.), their clients, compute infrastructure nodes (worker nodes interacting with remote storages) and data caches (XCache, etc.) that exchange large amounts of data over wide area network.
- Collectors (packet and flow label collectors): entities that process network traffic and collect flow identifiers, either directly from packets or by listening to UDP fireflies. Collectors are usually implemented in hardware, i.e. they require specific capability in the network equipment to either mirror the traffic and process it offline or directly capture and compute statistics in-line. In the packet marking case, the needed capability is to read and process IPv6 headers. In the flow labelling case, this can be performed either by a dedicated server or a network of servers (receivers) or by advanced network processing systems that can extract and analyse traffic offline, such as ESnet's high-touch service [2].
- Data management systems (data transfer orchestrator): Any entity that processes data management requests on behalf of the community and participates in the data management

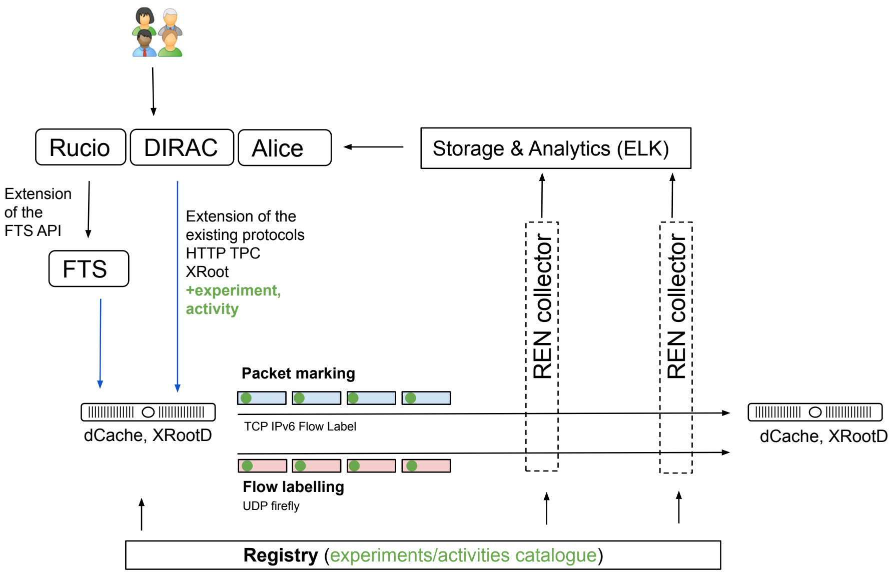

Figure 1: Scitags framework architecture for the WLCG use case showing interaction between data transfer agents (XRootD, dCache), Research and Education network (REN) collectors, data management agents (Rucio, DIRAC, Alice O2) and Scitags registry. The figure shows both packet marking and flow labelling exchanges between the data transfer agents.

workflow by propagating the community and its activity encoding all the way to *data transfer agents*. In WLCG's case this would be experiment's data management systems and transfer orchestrators such as Rucio, DIRAC, Alice O2, FTS, etc.

- Registry: a centralised component that stores the mapping between experiments, activities, and their numerical encoding, which are used by data management agents to encode the experiment and activity in the network traffic and collectors that can decode it.
- Storage and Analytics: Storage and analytics facilities hosted by network providers that can integrate R&E specific data, generate a global view, provide API access, and perform analytics. This can be used to generate information on how the network is being used, identify potential problems, and improve performance. Network providers are also globally interconnected, which means they can provide a global view of research and education activity. This can be used to track trends, identify collaborations, and discover new opportunities.

A typical life cycle of the Scitag *flow identifier* starts in the data management systems, where it is introduced and propagated through the experiment data management workflow via specific protocols (such as HTTP-TPC, Xroot, etc.) or API calls until it reaches the data transfer agent. The data transfer agent's main role is to introduce the Scitag identifier via packet marking or flow labelling on the network, so that collectors can capture and process the information (e.g. to correlate with traffic usage of the network flows). The registry is used by the data management systems to encode the community and its activity (numerical encoding needed for packet marking and flow labelling), and then later on the collectors use the registry for decoding these values.

More information on the framework with detailed technical specifications and examples can be found in the Scitags technical specification [3]. Members of the initiative have also published an IETF draft on packet marking using the IPv6 flow label that can be found in [4].

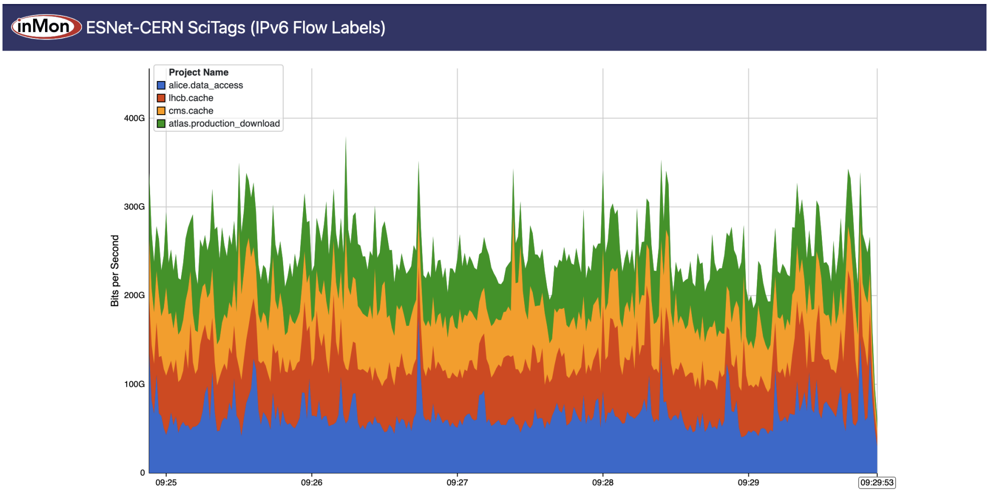

Figure 2: Scitags live demonstration at SC22 showing aggregated traffic between storage endpoints performing packet marking as seen by the network provider; fractions of the total usage by the experiments and their activities is shown in different colours [5].

3 Implementation Status

The implementation of the Scitags framework and, in particular, its WLCG use case, has made significant progress. A feasibility study was conducted as part of the WLCG Data Challenge in 2022 (DC22) in collaboration with ESnet and the AGLT2, KIT, BNL, and UNL WLCG sites. This study demonstrated the feasibility of correlating the *flow labelling* information sent from data transfer agents with traffic information seen by network providers (via netflow). Since *flow labelling* is a novel way to identify network flows, this step has been critical to ensure it works and can be used by network providers.

Scitag *Packet marking* at scale was demonstrated at SuperComputing 2022 (SC22), with packet marking running between two storage systems running XRootD at 200 Gbps, see Fig. 2. One of the challenges in packet marking is to ensure that it can be broadly deployed by various different data transfer agents (including those that don't have access to Linux kernel APIs) and therefore it was important to showcase such implementation as well as its performance.

Flow labelling implementation in storage systems also made progress, with full support in XRootD 5.0 and a prototype implementation in dCache that is currently being tested. Additionally, a dedicated component called *flowd*, which implements both flow labelling and packet marking for any third-party storage system, has been developed and used for experimental tests during DC22 and SC22 [6].

The registry is now in production, and preparatory work has started in data management systems (Rucio, FTS) to extend the existing transfer protocols in order to propagate the flow identifiers to the data transfer agents.

4 Related Work

A review of various different approaches for packet marking and flow labelling that we considered can be found in [7]. It summarises different technical approaches in flow labelling and packet marking such as IPv6 addressing, IPv4 options, IPv6 extension headers, SRV6, MPLS and others, as well as their trade-offs. A comprehensive review of related work focused on packet marking can be found in the IETF RFC draft [4], which also includes feedback received from the IETF community. The initiative also closely follows the work in HEP's network orchestration and traffic engineering projects, such as GNA-G and SENSE/Autogole [8][9]. Although the main focus of Scitags is to improve network visibility, there is also interest in exploring potential ways to collaborate and find complementary areas of future work in the area of traffic engineering [10].

5 Plans and Evolution

The Scitags initiative is a long-term effort to improve the visibility and management of network traffic in data-intensive science. Plans for the near- to mid-term future include:

- During the upcoming SuperComputing 2023, demonstrate capabilities of collectors deployed by the network providers and demonstrate packet marking at scale on 400Gbps infrastructure.
- Showcase the flow labelling and packet marking of Scitags from production storage(s) to collectors (network segments) during the WLCG Data Challenge 2024. Monitor what fraction of the traffic can be identified and correlated with the network flows seen by the network providers.
- Engage other scientific communities, network providers, and data management systems to adopt Scitags and improve the visibility and management of network traffic for data-intensive sciences.
- Collaborate with R&E network providers to develop collectors capable of capturing fireflies, reading, and accounting for marked packets and able to be deployed by other R&E networks. For example, Jisc is now running a collector and receiving fireflies from five UK sites.
- Develop and deploy easy-to-use collectors to gather regional Scitags information and be capable of forwarding data to other regional or global collectors.
- Define, document, and prototype analysis tools and storage infrastructures that allow Scitags data to be easily correlated with existing flow data and effectively displayed and organised for researchers and network engineers to use.
- After presenting and discussing the packet marking solution at the IETF117 meeting to explore further the potential to encode packet marking in the IPv6 hop-by-hop Extension Headers as an alternative to the Flow Label.

6 Conclusion

The efforts of the Research Networking Technical Working Group and the Scitags initiative have created a set of tools, standards, and proof-of-concept demonstrators that show the feasibility of identifying the owner (community) and purpose (activity) of the research and education network traffic anywhere in the network. While there are still significant areas of work required to move these capabilities into production for the broader community, we are confident that we will have enough data transfers being marked and identified to validate the approach in time for the WLCG Data Challenge 2024. We also plan to continue to improve the tools and components to make them usable by researchers outside of high-energy physics, ultimately leading to better understood, organised, and managed global research data transfers.

7 Acknowledgements

We want to explicitly acknowledge the support of the National Science Foundation, which supported this work through OSG: NSF MPS-1148698 and IRIS-HEP: NSF OAC-1836650 grants. Furthermore, we acknowledge our collaborations with the CERN IT, WLCG, WLCG DOMA, and LHCONE/LHCOPN communities, who also participated in this effort. Figure 2 showing the live demonstration from SC22 is used with permission of Neil McKee from InMon Corp.

References

[1] E. Martelli, S. Stancu, Journal of Physics: Conference Series **664**, 052025 (2015)

[2] Z. Liu, B. Mah, Y. Kumar, C. Guok, R. Cziva, *Programmable Per-Packet Network Telemetry: From Wire to Kafka at Scale*, in *Proceedings of the 2021 on Systems and Network Telemetry and Analytics* (Association for Computing Machinery, New York, NY, USA, 2021), SNTA '21, p. 33–36, ISBN 9781450383868, https://doi.org/10.1145/3452411.3464443

[3] *Flow and Packet Marking Technical Specification*, https://docs.google.com/document/d/1x9JsZ7iTj44Ta06IHdkwpv5Q2u4U2QGLWnUeN2Zf5ts/edit?usp=sharing (2023), [Online; accessed 22-August-2023]

[4] D.W. Carder, T. Chown, S. McKee, M. Babik, Internet-Draft draft-cc-v6ops-wlcg-flow-label-marking-02, Internet Engineering Task Force (2023), work in Progress, https://datatracker.ietf.org/doc/draft-cc-v6ops-wlcg-flow-label-marking/02/

[5] *sFlow: Scientific network tags (scitags)*, https://blog.sflow.com/2022/11/scientific-network-tags-scitags.html (2023), [Online; accessed 22-August-2023]

[6] M. Babik, T. Sullivan, *flowd* (2023), https://github.com/scitags/flowd

[7] *RNTWG Packet Marking Review*, https://docs.google.com/document/d/1aAnsujpZnxn3oIUL9JZxcw0ZpoJNVXkHp-Yo5oj-B8U/edit?usp=sharing (2023), [Online; accessed 22-August-2023]

[8] Experience, ed., *NRE-016: AutoGOLE/SENSE: End-to-End Network Services and Workflow Integration*, Vol. 2021 (Supercomputing Conference 2021 (SC21), St. Louis, MO, 2021), https://sc21.supercomputing.org/app/uploads/2021/11/SC21-NRE-016.pdf

[9] *Global Network Advancement Group*, https://www.gna-g.net/ (2023), [Online; accessed 22-August-2023]

[10] C. Misa Moreira, E. Martelli, T. Cass, EPJ Web Conf. **to appear** (2024)