Scitags: A Standardized Framework for Traffic Identification and Network Visibility in Data-Intensive Research Infrastructures

Garhan Attebury[8], *Marian* Babik[2], *Dale* Carder[3], *Tim* Chown[6], *Pablo* Collado[11], *Andrew* Hanushevsky[4], *Bruno* Hoeft[7], *Andrew* Lake[3], *Michael* Lambert[10], *Shawn* McKee[1], *Karl* Newell[9], and *Tristan* Sullivan[5]
for the Scitags Initiative*

[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 Nebraska, Lincoln, NE, USA
[9]Internet2, Ann Arbor, MI, USA
[10]Pittsburgh Supercomputing Center, Carnegie Mellon University, Pittsburgh, PA, USA
[11]Univerisdad Autónoma de Madrid, Spain

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 an update on the status of the work carried out to tackle this challenge by 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 2026. The Scitags initiative is a long-term effort to improve the visibility and management of network traffic for data-intensive sciences. It has 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 to interconnect sites, data centres, and instruments globally. This network traffic uses purpose-built networks (such as LHCOPN and LHCONE) and general research and education (R&E) networks [1]. The management of such networks is becoming increasingly complex due to the growing number of science projects utilizing the same infrastructure, the increasing volume of data being transferred,

*Acknowledgments: This work has been supported by OSG: NSF MPS-1148698 and IRIS-HEP: NSF OAC-1836650 grants. Furthermore, we acknowledge our collaborations with the CERN IT, WLCG project and experiments and LHCONE/LHCOPN communities, who also participated in this effort. More details at:www.scitags.org.

and the diverse requirements and technologies employed. One of the challenges is the lack of visibility into which scientific communities are generating specific network flows and their intended purpose. Without sufficient visibility, it becomes difficult to understand network utilization, optimize performance, and troubleshoot operational issues.

To address this challenge, the Scitags initiative was created to develop a generic framework and standards for identifying the owner (community) and associated activity of the network traffic. The goal is to implement this model in production for any global user of R&E networks, starting with the High-Energy Physics (HEP) experiments and progressively expanding to other data-intensive sciences.

This paper outlines the project's current state, detailing the evolving framework, tools, and plans for future evolution.

2 Framework

The Scitags initiative has developed a framework to help identify communities and their activities at the network level. In the context of the Worldwide LHC Computing Grid (WLCG), this involves the identification of experiments, such as ATLAS and CMS, and their associated activities, such as production, analysis, data challenge, etc. The framework is engineered to deliver high-fidelity visibility for data-intensive scientific communities. It is open source and available to any interested community. The framework promotes quantification of global behaviour, flexibility, and scalability through the decoupling of data producers and consumers and compatibility with the existing network monitoring tools run by R&E network providers.

Two mechanisms are proposed to identify the community and activity related to network traffic:

- *Packet marking:* This process involves embedding a community and activity within the header of each network packet, thereby indicating the community and activity to which the traffic belongs. This is an in-band approach, as the *scitag* (community/activity) is embedded directly in the original data transfer.
- *Flow labelling:* This approach uses a separate communication channel to convey the *scitags* and associated metadata, allowing identification of reference data transfer. This is an out-of-band approach as we're using a separate stream of packets to carry *scitags* as well as a reference to the original transfer that they belong to.

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

- Storage: Perform the actual transfers of the HEP data sets and implement either *flow labelling* or *packet marking* to identify the traffic generated. Examples of storage that integrate *scitags* are XRootD, EOS, StoRM and dCache.
- Collectors: Process network traffic and gather flow identifiers at the site boundary or anywhere in the network, either directly from packets or by monitoring flow labelling streams. These are, for example, Site and/or R&E collectors.
- Data management systems: These process data management requests and propagate community and activity encoding to storages, e.g., Rucio, DIRAC, FTS.
- Registry: This centralized component provides a catalogue of the supported communities and activities as well as their numerical encoding.
- Storage and Analytics: These facilities, hosted by network providers or sites, integrate network specific data with *scitags*, generate a global view, provide API access, and perform analytics.

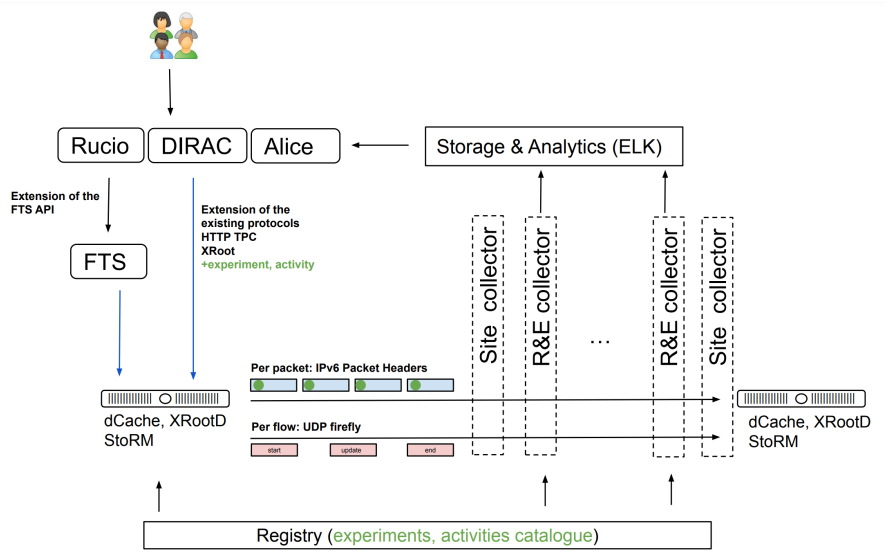

Figure 1: Scitags framework architecture for the WLCG use case showing interaction between storages (XRootD, EOS, StoRM), Site or Research and Education network (REN) collectors, Data Management Systems (Rucio, DIRAC, Alice O2) and Scitags registry. The figure shows both *packet marking* and *flow labelling* exchanges between the data transfer agents.

The *scitag* lifecycle begins in data management systems, where it is introduced and propagated through the data management workflow. Subsequently, the storages use *packet marking* or *flow labelling* to disseminate the identifier on the network, enabling collectors to capture and process the information. The registry is used to ensure that common encodings of *scitags* are used across all elements of the architecture.

More information on the framework can be found in the Scitags technical specification [2] and white paper [3].

3 Packet Marking

Packet marking is a method used by Scitags to identify the community and activity associated with network traffic by adding a *scitag* to every network packet sent.

To tag a packet, a suitable field must be chosen in the packet header. In the initial study by the Research Networking Technical WG [4], it was determined that no suitable field exists in the IPv4 header. Because the vast majority of WLCG traffic is IPv6, the focus turned to marking IPv6 packets.

Two ways of marking IPv6 packets are under investigation:

- Using the *IPv6 flow label field*: Overwriting the flow label field is a lightweight operation compared to inserting a new header into a packet, and thus the performance impact of packet marking using the flow label is expected to be minor. However, the size of the flow label field is limited to 20 bits, which limits the potential number of communities and activities that can be encoded.
- Using a *destination options extension header*: This is an optional header in IPv6 that carries information that should only be examined by the destination node(s). It is used to send

additional control information to the destination and can hold a much larger amount of information.

Using an IPv6 *hop-by-hop extension header* was considered as well, but quickly rejected as a non-viable option. The presence of a hop-by-hop header in a packet forces routers to analyse that packet using the slow path, and in testing it was found that most routers simply prefer to drop such packets.

The *packet marking* can be performed by a stand-alone software that is intended to be run on the storages. When a flow is to be marked, it receives a signal identifying the flow and the marking to be applied. The actual *packet marking* implementation relies on the extended Berkley Packet Filter (eBPF), which is a capability of the Linux kernel that allows user code to be injected into the kernel at runtime. The eBPF code runs when a specific kernel action is taken, in our case when a packet is sent to the network interface. Depending on the marking strategy chosen, either the flow label in the packet header is overwritten or an extension header is inserted into the packet.

More details on the related work and how community and activity is encoded in the headers can be found in the Scitags IETF draft on packet marking [5].

4 Flow Labelling

Flow labelling is a mechanism used within the Scitags framework to identify the community and activity associated with network traffic. Instead of modifying every packet, flow labelling uses a separate communication channel to carry both the *scitag* and metadata to identify the reference data transfer. Flow labelling is achieved via specially crafted UDP packets, called *fireflies*. Fireflies have syslog format with a defined, versioned JSON schema and the following characteristics:

- Content: Apart from community/activity, the UDP packet also carries metadata of the original transfer, i.e. protocol, source/destination addresses and source/destination ports.
- Destination: Packets are sent to the same destination (port 10514) as the data transfer they are labelling, and these packets are intended to be world readable. Packets can also be sent to a specific regional or global collectors.
- Syslog Format: The use of syslog format facilitates processing of the packets by Logstash or similar receivers.
- Compatibility: Fireflies work for both IPv4 and IPv6 and the length of their content is not limited as long as it fits within a single frame. This makes it possible for fireflies to carry additional socket information such as bytes sent/received, Round-Trip Time (RTT), congestion algorithm, etc.

More details on the schema and examples of fireflies can found in the Scitags technical specification [2].

5 Packet Marking and Flow Labelling Service

Storages such as XRootD, EOS, dCache and StoRM deal with a large amount of complexity and are, in general, developed in different programming languages. If each of them were to implement the needed functionality for *packet marking* and *flow labelling* the impacted communities would duplicate their efforts. In addition, as packet marking relies on the eBPF capability and Linux kernel interactions, some storages lack the required interfaces and require that such functionality is off-loaded to a third-party service.

Flowd [6] and flowd-go [7] are software implementations that solve this issue by providing a flexible architecture, allowing the storage services to implicitly *mark packets* and *label*

flows with minimal changes to their current implementations. The ease of integration is a consequence of the plugin design shown in figure 2. Storages need only leverage one of the available plugins or propose a new one to gain access to flowd's packet marking capabilities. The different marking strategies are implemented as backends. Flowd and flowd-go are intended to run as services alongside the storage systems.

The initial implementation was flowd, which was written in Python, and has been used to test various approaches and methods for packet marking and flow labelling. Given the large data rates that must be handled, a new implementation in Go, flowd-go, was developed in an effort to increase packet marking performance. Flowd-go is still under development and it is expected to become the standard implementation.

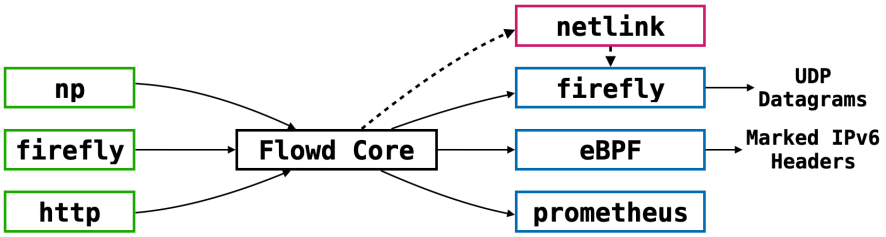

Figure 2: Flowd's architecture. Plugins are represented by green boxes and backends are shown in blue. Given its privileged position alongside storages, flowd can leverage Linux's Netlink [8] subsystem (red) to enrich outgoing fireflies with socket-level statistics.

6 Data Challenges and Demonstrations

6.1 Supercomputing 2024

Two methods of *packet marking* were compared during a network demonstration at the Supercomputing 24 conference. A 400 Gbps network path was provisioned between the University of Victoria and the exhibition area in Atlanta, with a 400 Gbps-capable server at each end.

Memory-to-memory transfers were generated using sixteen parallel iperf3 streams. The Scitags demo ran for two hours each day of the conference. For forty minutes, no packet marking was enabled; then IPv6 flow label marking was enabled for forty minutes; and for the final forty minutes, IPv6 destinations option extension header marking was enabled. Each iperf3 stream was given its own random pair of experiment and activity labels. The packet marking was performed by the UVic server.

The network bandwidth, CPU usage, and memory usage on the UVic server were monitored using Prometheus. Figure 3 shows the results from the first day of the demo. The main parameter of interest is the impact on bandwidth of the two different packet marking methods. The average bandwidth achieved was approximately 260 Gbps and 240 Gbps with no marking and IPv6 flow label marking, respectively. With destination option header marking, the bandwidth was initially around 220 Gbps, before falling sharply to 150 Gbps around 11:30 and slowly increasing again. Around 11:50, it fell sharply again and slowly increased.

In an attempt to understand this behaviour, the bandwidth was limited to 192 Gbps on the second day of the demo. The Prometheus dashboard is shown in Figure 4. When packet marking was disabled or IPv6 flow label marking was used, the bandwidth target of 192 Gbps was achieved for the duration of the transfer. Destination option header marking continued to suffer from the same issue as the first day; the bandwidth fell steeply before slowly increasing three times.

Figure 3: Prometheus monitoring dashboard for the first day of the demo. The values shown are from the UVic server. In the top left, bandwidth sent and received are plotted; only bandwidth sent is visible on this scale. In the bottom right, used and available memory are shown; the impact of memory usage from either packet marking method is invisible on this scale. The other four panels show the CPU usage for the four NUMA nodes on the server. The NIC was on NUMA node three, as can be seen from the higher utilization of it compared to the other three NUMA nodes.

Figure 4: Prometheus monitoring dashboard for the second day of the demo, when the bandwidth was capped at 192 Gbps. Nevertheless, the extension header marking still shows the same instability as it did on day one.

To demonstrate the ability to read the *scitags* applied to the packets, sFlow data was exported from the server at UVic to a collector machine at SC24 [9]. This machine hosted a webpage showing the experiment and activity labels from traffic in the last hour. Figure 5 shows the results at the end of the first day. This is an important aspect of the demo, as the ability to make plots like this on any network link is one of the main goals of the project.

6.2 WLCG Data Challenge 2024

WLCG Data Challenge 24 (DC24) involved the first pilot deployment of Scitags and demonstrated its potential for network usage analysis. Around 80% of the CERN CMS storage and

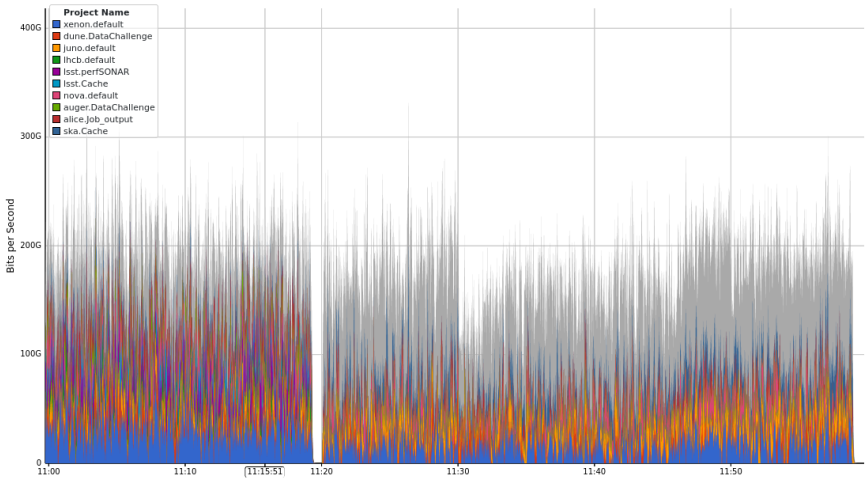

Figure 5: Bandwidth sent by the UVic server, grouped by the packet labels. The data from the last hour is plotted; thus, IPv6 flow label marking was used for about the first twenty minutes shown, and extension headers marking was used for the remaining time period. The bandwidth drop when using extension headers is visible. Each iperf3 stream received a randomly chosen experiment and activity pair.

the entire production storage at the University of Nebraska utilized Scitags deployment with *flow labelling* functionality enabled. The deployment successfully demonstrated end-to-end *scitag* propagation from data management systems to storage facilities for both ATLAS and CMS experiments. It also verified that XRootD and EOS storage systems could transmit fireflies, which were then collected and displayed in real-time on ESnet's monitoring dashboard [10]. Some of the highlights that were captured include:

- Flow count: For both sites, we were able to plot the overall count of network flows split by experiment and activity, cf. Fig. 6.
- Flow durations: Maximum duration of flows was calculated from the respective start and end times reported in the fireflies and then split by experiment/activity. Overall, during the two weeks of the DC24 we have observed network flow durations not exceeding one minute, which isn't representative of the usual production traffic patterns where duration is in hours.
- Non-FTS Traffic Split by Applications: University of Nebraska (UNL) also tracked traffic not coming from FTS, which revealed that during DC24 there was also significant traffic generated by direct remote access from the worker nodes at the UNL compute facility. This has highlighted the total amount of traffic hitting wide-area network directly from the worker nodes at UNL as well as underlying applications, both of which are currently not tracked by any other monitoring system.

7 Implementation Status

The implementation of the Scitags framework, in particular its WLCG use case, has made significant progress. A pilot production deployment was conducted as part of the WLCG Data Challenge in 2024 (DC24) in collaboration with ESnet, CERN and UNL to showcase *flow labelling* capabilities. Scitags *packet marking* at scale was demonstrated at SuperComputing 2024 (SC24) using different strategies.

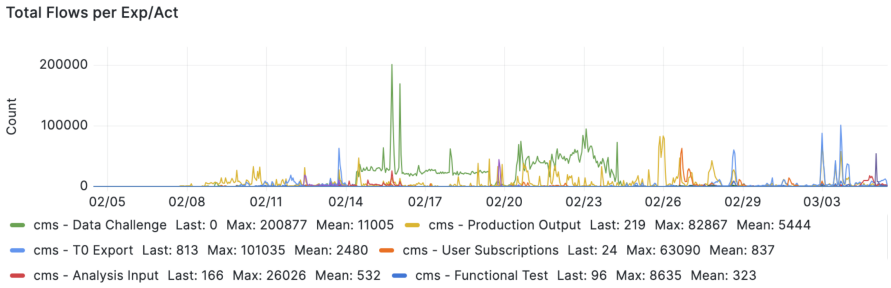

Figure 6: Total number of network flows for CERN EOS CMS storage, categorized by activity. The green line specifically highlights the period of the WLCG Data Challenge 2024, which ran from February 12th to 23rd.

Flow labelling implementation in storage systems has made significant progress, with full support in XRootD 5.0+, StoRM 1.4.3, EOS 5.2.19 and an initial release in dCache 10. Additionally, flow marking and labelling services were updated with new features and bugfixes, with *flowd* release 1.1.6, which implements a number of enhancements to integrate *scitags* capabilities in Kubernetes environments. *flowd-go* version 2.0 has been released with support for production workflows for XRootD, EOS and StoRM.

Data management systems have implemented support for Scitags, Rucio in 32.4.0 and FTS/GFAL in versions 3.2.10/2.21.0. Both have been deployed and tested in production during the WLCG Data Challenge 2024 for both ATLAS and CMS. Alice data management systems also support Scitags in production for the entire Alice infrastructure. We have also engaged with LHCb and plans were made to add Scitags support in DIRAC.

The registry has been in production since 2023 and additional communities and activities have been registered and the overall deployment workflow has been improved. Due to its comprehensive catalogue of research and education network users and their activities, the system has proven valuable for additional applications, such as LHCONE's multi-BGP communities, where it helps map research groups and experiments to their corresponding BGP communities. [11].

8 Conclusion

The Scitags initiative is a long-term effort focused on enhancing the visibility and management of network traffic within data-intensive scientific research. By developing a standardized framework for identifying the owner (community) and purpose (activity) of network traffic, Scitags addresses the challenges faced by High-Energy Physics (HEP) and other data-intensive sciences in understanding and optimizing their global network flows.

Significant progress has been made in implementing the Scitags framework, with support in various storage systems and data management tools. Pilot production deployments, such as the WLCG Data Challenge 24 (DC24), have demonstrated the potential of Scitags for network usage analysis.

Looking ahead, the plan is to prepare for wider production deployment before WLCG Data Challenge in 2027 where we plan to demonstrate the benefits in correlating Scitags information with other R&E network monitoring systems, such as ESnet High-touch service [12]. Future demonstrations at SC25 will focus on evaluating Flowd-go performance for packet marking. We also plan to engage with other scientific communities, integrate additional data management systems and storages and work with other R&E networks to expand the adoption and impact of the framework.

References

[1] E. Martelli, EPJ Web of Conferences **295** (2024)

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

[3] G. Attebury, M. Babik, D. Carder, T. Chown, A. Hanushevsky, B. Hoeft, A. Lake, M. Lambert, J. Letts, S. McKee et al., EPJ Web of Conf. **295**, 01036 (2024)

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

[5] 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/`

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

[7] M. Babik, T. Sullivan, P. Collado, *flowd-go* (2024), `https://github.com/scitags/flowd-go`

[8] T. kernel development community, *Introduction to Netlink*, `https://www.kernel.org/doc/html/latest/userspace-api/netlink/intro.html` (2023), [Online; accessed 15-February-2025]

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

[10] *ESnet: Scientifc Network Tags Live Grafana Dashboard*, `https://dashboard.stardust.es.net/d/b8dddac0-5b24-4739-9c8d-e88a05c1344f` (2024), [Online; accessed 19-February-2025]

[11] E. Martelli, T. Cass, To appear in EPJ Web of Conf. (2025)

[12] 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. 3336, ISBN 9781450383868, `https://doi.org/10.1145/3452411.3464443`