

Research Networking Technologies Working Group

Shawn McKee (University of Michigan Physics), [Marian Babik \(CERN\)](#), Tim Chown (Jisc),
Andrew Hanushevsky (SLAC National Accelerator Laboratory), Andy Lake (ESnet),
Tristan Sullivan (University of Victoria), Bruno Hoefft (Karlsruhe Institute of Technology (KIT)),
James Letts (University of California, San Diego), Dale Carder (ESnet), Garhan Attebury (UNL),
Michael Lambert (Pittsburgh Supercomputing Center), Joe Mambretti (Northwestern University),
Karl Newell (Internet2), et al. - on behalf of the RNT Working Group and Scitags project

LHCOPN/LHCONE #54 meeting

March 18-20, 2025

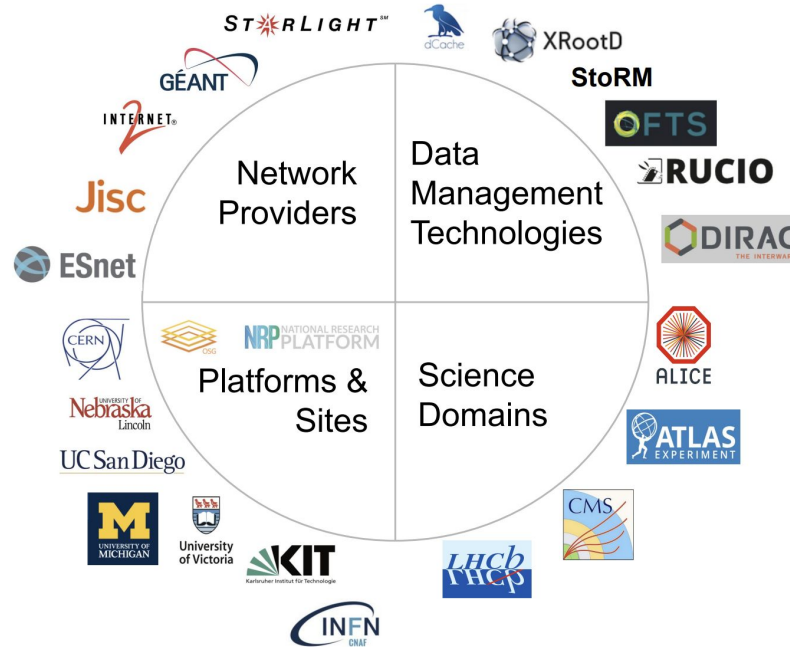
Research Networking Technologies WG

- Started in 2020 based on [Network Functions Virtualisation White Paper](#)
 - Overview of networking technologies for both LAN/WAN at that time
 - Highlighted focused areas for R&D with potential impact to HEP networking
- WAN topics:
 - **Network Visibility**
 - **Scitags project** (Packet marking/Flow labelling)
 - This talk; and also Packet Marking Demo (Tristan)
 - MultiOne project (Edoardo)
 - **Network Performance**
 - Packet pacing; jumbo frames, TCP congestion algos (BBRv3), etc.
 - Results of EOS Jumbo frames testing (Maria)
 - **Network Orchestration**
 - GNA-G, AutoGOLE/SENSE project, GRP

Scitags Project

scitags.org

Network Flow and Packet Marking for
Global Scientific Computing



Network Visibility and Scitags

- **Scientific Network Tags** (scitags) is an initiative promoting identification of the science domains and their high-level activities at the network level.

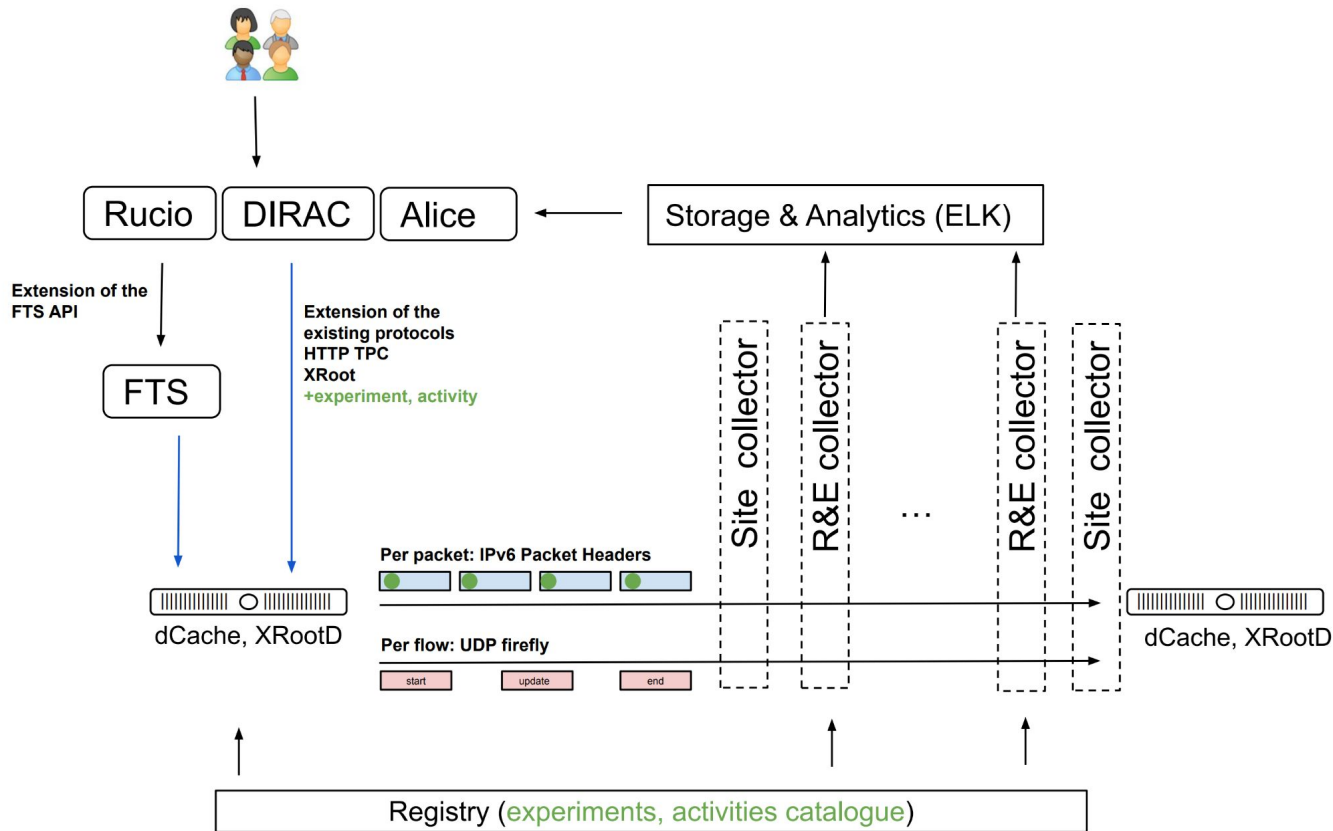


- **Experiments/Collaborations** can better understand how their network flows perform along the path
 - Network performance studies, sites profiling
- **Research and Education Network Providers** gain visibility into sources and purpose of the network traffic
 - Enable **tracking** and **correlation** of network flows with REN systems
 - Facilitates debugging and capacity planning

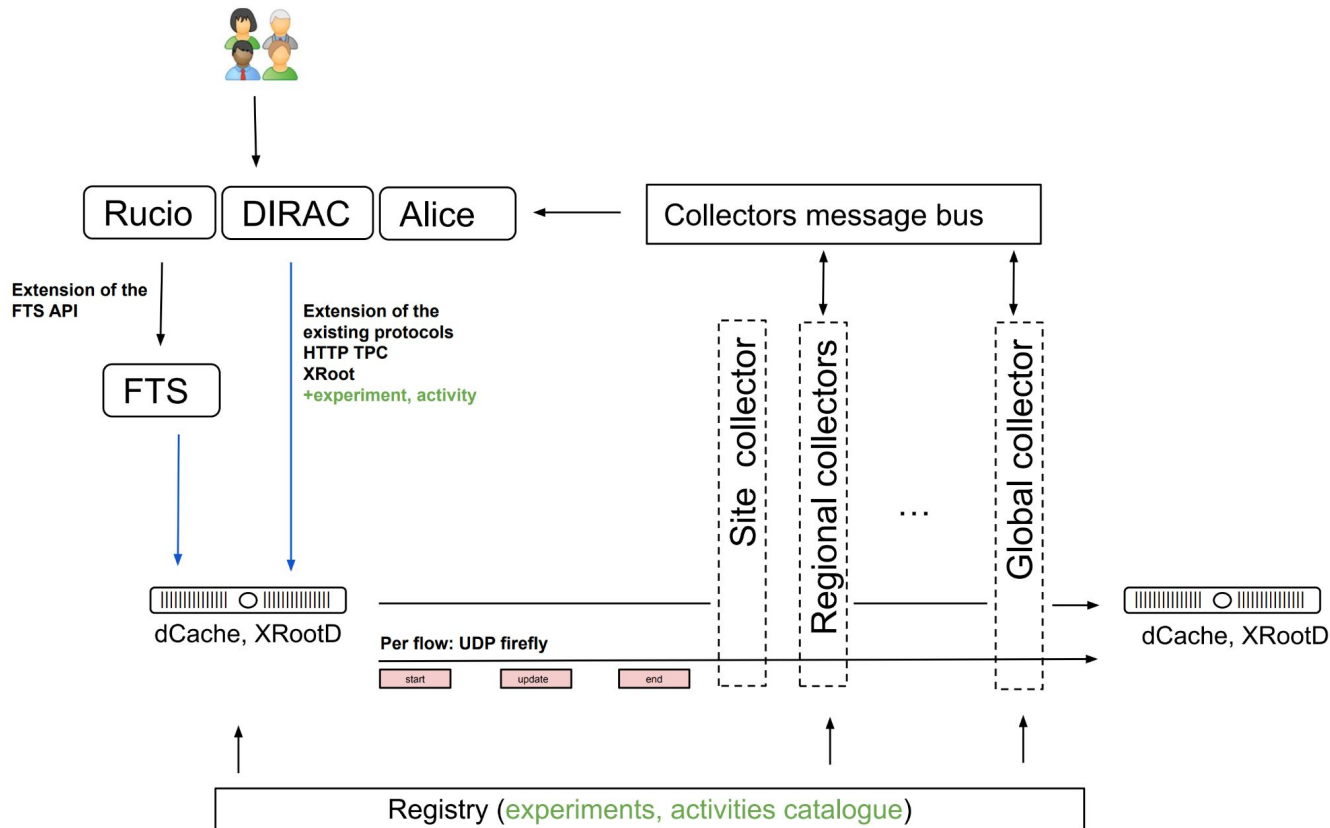
Scitags Framework Rationale

- **Open platform** to be used by any data-intensive science community
- **Identify the owner (experiment) and purpose (activity) of the traffic**
- Define a **standard(s)** for exchange of information between scientific communities, sites and network operators
 - Packet marking - encoding exp/activity directly in packets
 - Flow labeling - sending a separate UDP packet (firefly) with metadata
- **Enable tracking and correlation with existing network flow monitoring and/or existing monitoring systems deployed by R&E networks**
- Quantify global behaviour and analyse trade-offs at scale

How scitags work



How scitags work



Finding More Information: <https://scitags.org>

Code

Technical Spec

Mailing List

scitags.org

Network Flow and Packet Marking for
Global Scientific Computing



Scientific network tags (scitags) is an initiative promoting identification of the science domains and their high-level activities at the network level.

It provides an open system using open source technologies that helps *Research and Education (R&E) providers* in understanding how their networks are being utilised while at the same time providing feedback to the *scientific community* on what network flows and patterns are critical for their computing.

Our approach is based on a network tagging mechanism that marks network packets and/or network flows using the science domain and activity fields. These tags can then be captured by the *R&E providers* and correlated with their existing netflow data to better understand existing network patterns, estimate network usage and track activities.

The initiative offers an **open collaboration on the research and development of the packet and flow marking prototypes** and works in close collaboration with the scientific storage and transfer providers to enable the marking capability. The project is currently in the prototyping phase and is open for participation from any science domain that require or anticipate to require high throughput computing as well as any interested *R&E providers*.

Participants



Upcoming and Past Events

- March 2022: LHCOPN/LHCONE workshop
- November 2021: GridPP Technical Seminar (slides)
- November 2021: ATLAS ADC Technical Coordination Board
- October 2021: LHCOPN/LHCONE workshop (slides)
- September 2021: 2nd Global Research Platform Workshop (slides)

Presentations

Current Status: Experiments

CMS

- **Fireflies enabled in production at**
 - CERN, UCSD, UNL, Caltech, Wisconsin, Purdue, Florida, UK (Brunel, QMUL)
- **There is a CMS-wide campaign underway to deploy fireflies**, with agreement to enable them at all CMS sites running xrootd/EOS.
- Flow labelling from compute/worker nodes is in progress.
 - Agreed with CMSSW developers to update relevant components

ATLAS

- Fireflies enabled at CERN EOS ATLAS production since January
- An **ATLAS campaign is pending support in dCache**, which remains a major concern for ATLAS as many sites use it

Validations and Capability Data Challenges have been discussed with all WLCG experiments. This includes running ESnet High-Touch service telemetry correlated with fireflies.

Current Status: Experiments

ALICE

- Fireflies enabled at **CERN EOS ALICE production** since DC24
- Plans are open for wider deployment, mainly at ALICE US and UK sites.
- Interested in detailed TCP metrics for site profiling
- Site-level collectors and data aggregation at the site level has been discussed

LHCb

- **Enabling fireflies support in DIRAC** has been discussed
 - LHCb has agreed to review our requirements and specifications
- Enabling fireflies at all LHCb sites has been discussed, with dCache being a recurring concern.

SKA and Belle II

- Meetings are planned this week, tooling is ready
- Fireflies could be interesting for the network performance studies

Current Status: DDM & Storages

DDM:

- **Rucio** (from 32.4.0) and **FTS/gfal2** from 3.2.10/2.21.0
- Works for both **HTTP** and **XRoot** protocols (cf. technical specification)

Storages:

- **XRootD** provides [Scitags implementation](#) (from 5.0+)
 - In production at several sites
- **EOS** provides Scitags support from 5.2.19+
 - In production at CERN EOS ATLAS, CMS and ALICE
- **dCache** firefly support released in 10.0.2 (static config only)
 - Support for readout from XRoot and HTTP protocols underway
- [StoRM](#) provides Scitags support from 1.4.3+
 - In production at **CNAF** for all supported experiments

Collectors:

- Production deployments at [ESnet](#), Jisc, in discussion with GEANT
- Site-collector running at CERN

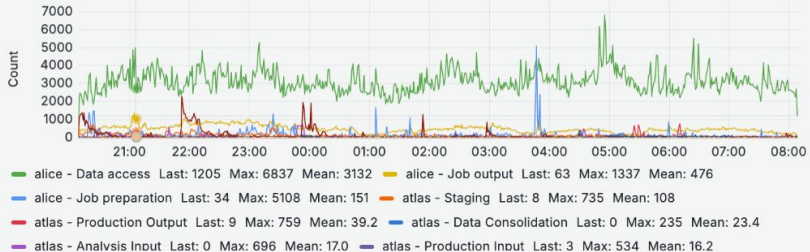
ESNet: Scitags Dashboard

Scientific Network Tags

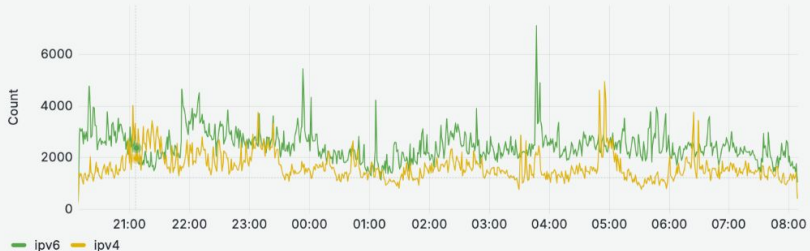
Menu: [Overview](#) | [Interfaces](#) | [Sites](#) | [Regionals](#) | [Transatlantic](#) | [LHCOPN](#) | [Scientific Network Tags](#)

This dashboard shows statistics related to flows marked with [Scientific Network Tag \(scitags\)](#) and sent to the ESnet firefly collector.

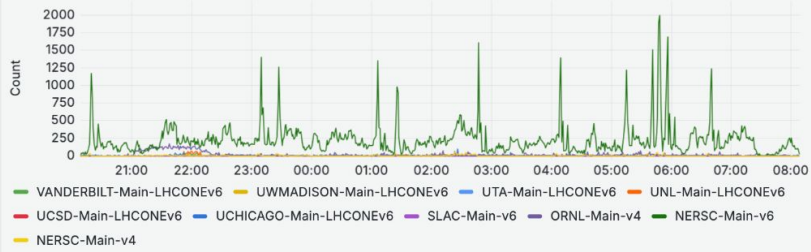
Total Flows per Exp/Act



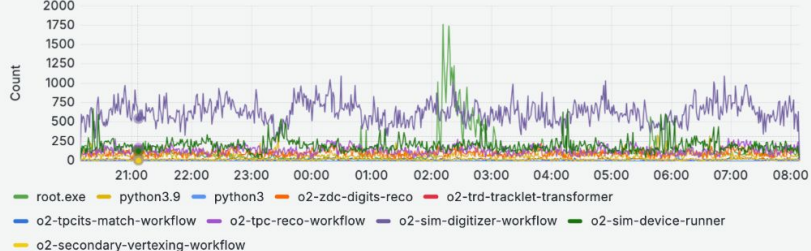
Total Flows per IP version



Total Flows per Prefix



Total Flows per Application (top10)

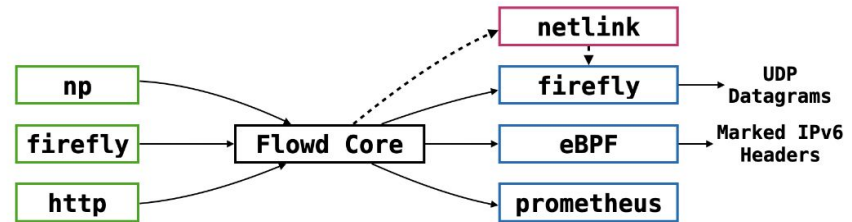


Packet Marking and Flow Labelling Service -flowd

Service and library to prototype and test various approaches to packet marking and flow labelling

Core features ([flowd-python](#)):

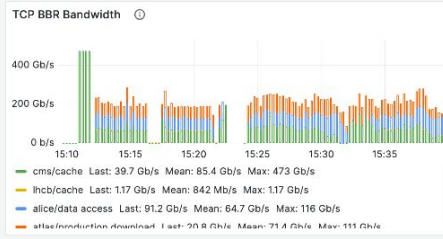
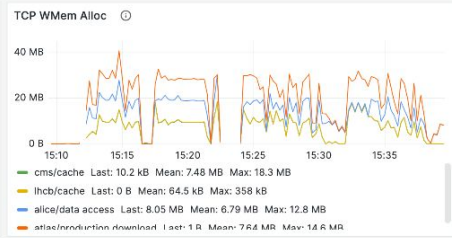
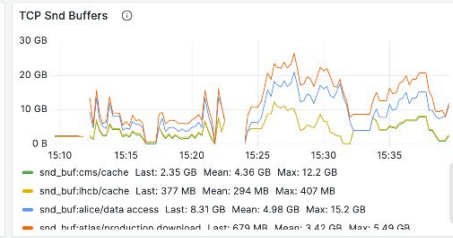
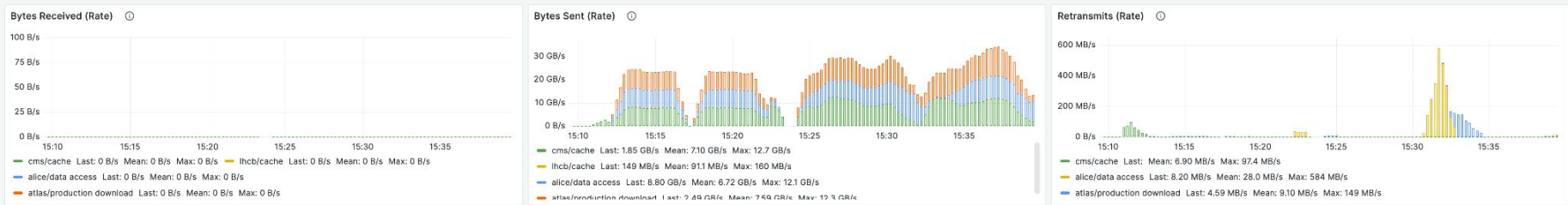
- Pluggable architecture
- Capability to work alongside storages
- Portable eBPF deployment
- Container & package distribution
 - Support for [Alma8](#), [Alma9](#), [RHEL8](#) and [RHEL9](#)
 - Also available as [Docker container](#)



Now also available in Go ([flowd-go](#)) - credits Pablo Soto from UAM

- Initial version, uses directly libBPF (supports multiple kernel versions)
- Extended to efficiently enrich fireflies with Linux TCP/IP stack information

Src: All Dest: All



Summary

- **Scitags (flow labelling) ready for production**
 - **R&Es can already receive UDP fireflies from multiple sources**
 - Significant progress has been made in enabling flow labelling within the LHC experiments (CMS, ATLAS, ALICE, LHCb)
 - Key areas of focus include production deployment, addressing storage system compatibility, and exploring the use of Scitags for data challenges and network performance analysis
 - We're **open to collaboration with other science domains** and R&E networks - please contact us if you have specific use cases/tools, etc.
- **Scitags (packet marking) - details in Tristan's talk**
 - We plan to demonstrate new and existing capabilities during WLCG mini-challenges, SC25 and DC27

Acknowledgements

We would like to thank the **WLCG**, **HEPiX**, **perfSONAR** and **OSG** organizations for their work on the topics presented.

In addition we want to explicitly acknowledge the support of the **National Science Foundation** which supported this work via:

- [OSG: NSF MPS-1148698](#)
- [IRIS-HEP: NSF OAC-1836650](#)

Backup slides

- **Flow Labeling** via **UDP packets (fireflies)**:
 - **Fireflies** are UDP packets in Syslog format with a defined, versioned JSON schema.
 - Packets are intended to be sent to the same destination (port 10514) as the flow they are labeling and these packets are intended to be world readable.
 - Use of syslog format makes it easy to send to Logstash or similar receivers.
 - Works for IPv4 and IPv6; content is not limited (as long as it fits in a single frame)
 - Apart from exp/act we now have also usage (bytes sent/rcv) and RTT in fireflies
 - “sFlow like” (unsampled) stream with additional metadata (science domain/activity)
- The detailed technical specifications are maintained on a [Google doc](#)
- The document also covers methods for communicating owner/activity and other services and frameworks that may be needed for implementation.

- **Fireflies collector is a basic software-based collector**
 - Processing incoming stream of UDP packets (fireflies)
 - **Logstash** plugins + filters
 - Participates in a network of collectors interconnected by a message bus
 - Uses **Kafka** as message bus to publish/subscribe
 - Optionally provides storage and analytics services
 - **ElasticSearch + Grafana**
 - Can enhance data from additional sources (sites, prefixes, ASNs)
- **Operational requirements**
 - Operated on best-effort basis
 - Can be VM, needs open access on port 10514
 - Logstash code and instructions available from Github

Registry

We have standardized the “experiment” and “activity” fields we use for both flow labeling and packet marking.

The **scitags.org** domain provides an API that can be consulted to get the standard values:

<https://api.scitags.org> or <https://www.scitags.org/api.json>

The underlying source of truth is a set of [Google sheets](#) that are maintained and writeable by a few stewards.

Note: the API provides the defined values **but** how the values are used in packet marking are specified in our [Google sheets](#) (bit location in IPv6 flow label)

```
{
  - experiments: [
    - {
      expName: "default",
      expId: 1,
      - activities: [
        - {
          activityName: "default",
          activityId: 1
        }
      ]
    },
    - {
      expName: "atlas",
      expId: 2,
      - activities: [
        - {
          activityName: "perfsonar",
          activityId: 2
        },
        - {
          activityName: "cache",
          activityId: 3
        },
        - {
          activityName: "datachallenge",
          activityId: 4
        },
        - {
          activityName: "default",
          activityId: 8
        },
        - {
          activityName: "analysis download",
          activityId: 9
        },
        - {
          activityName: "analysis download direct io",
          activityId: 10
        }
      ]
    }
  ]
}
```

Scitags Deployment

Site requirements:

- Run one of the supported storages (xrootd, EOS, StoRM)
- Configure scitags (xrootd example):

```
xrootd.pmark use firefly scitag  
xrootd.pmark domain any  
xrootd.pmark ffdest <scitags_collector>:10514
```

Collaboration requirements:

- Run one of the supported data management systems (Rucio, FTS)
- Activities registered in the Scitags registry

R&E network provider requirements:

- Deploy firefly collector (requires new/existing ELK stack)
 - Configure message bus export to global collector
- Configure Grafana dashboard
- Integrate with other information sources (sites, ASNs, prefixes, etc.)

Data Challenge 24

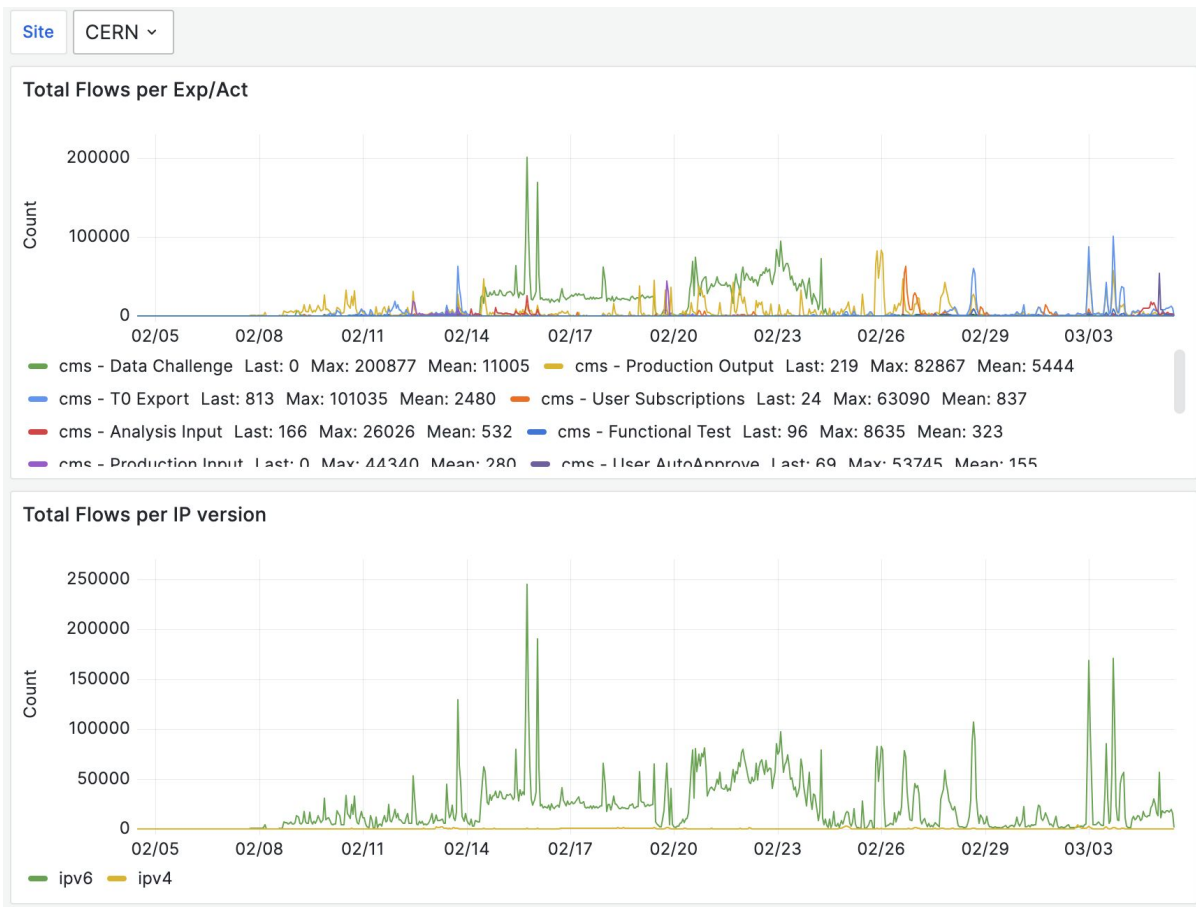
- **Scitags Deployment**

- 80% of EOS CMS (production), UNL production storage
- Flow labeling functionality (fireflies)

- **Results:**

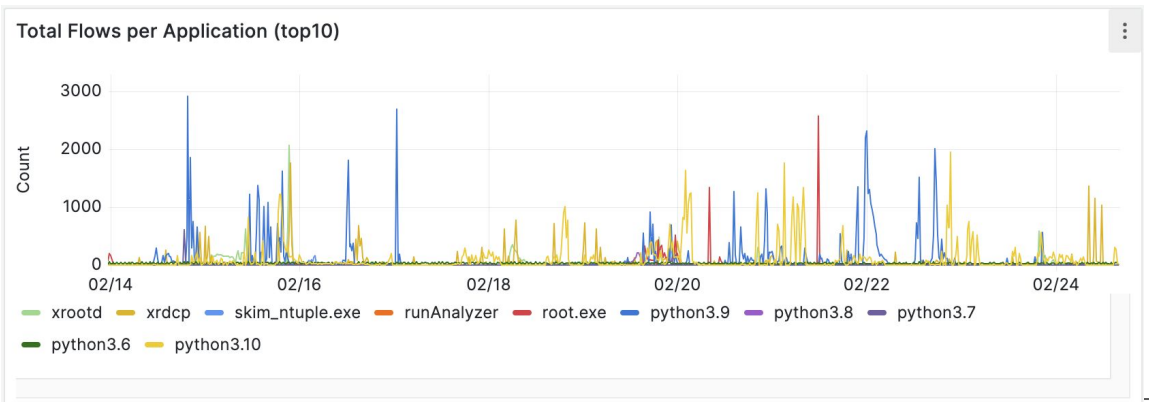
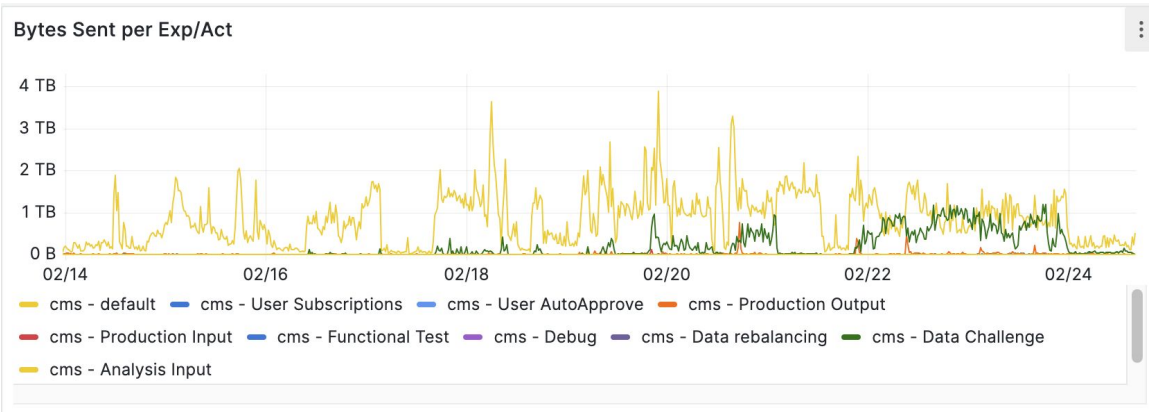
- **Confirmed the capability to propagate Scitags all the way to the storages (for both ATLAS and CMS)**
- Sending fireflies (from XRootd, EOS storages)
- **Collection and visualisation at ESnet collector**
 - Results shown in [live dashboard](#)
- With limited deployment we were able to get valuable insights into flow durations, their characteristics (splits by exp/activity), sources of IPv4 traffic (split by applications) and potential impact of new TCP congestion algorithms (performance correlated with flow data)

DC24: CERN EOS CMS plot showing split by experiment/activity and IP



DC24: University of Nebraska (UNL)

Bytes sent per Exp/Activity - shows non-FTS traffic was dominant
Capability to show a split by application (as reported by xrootd)

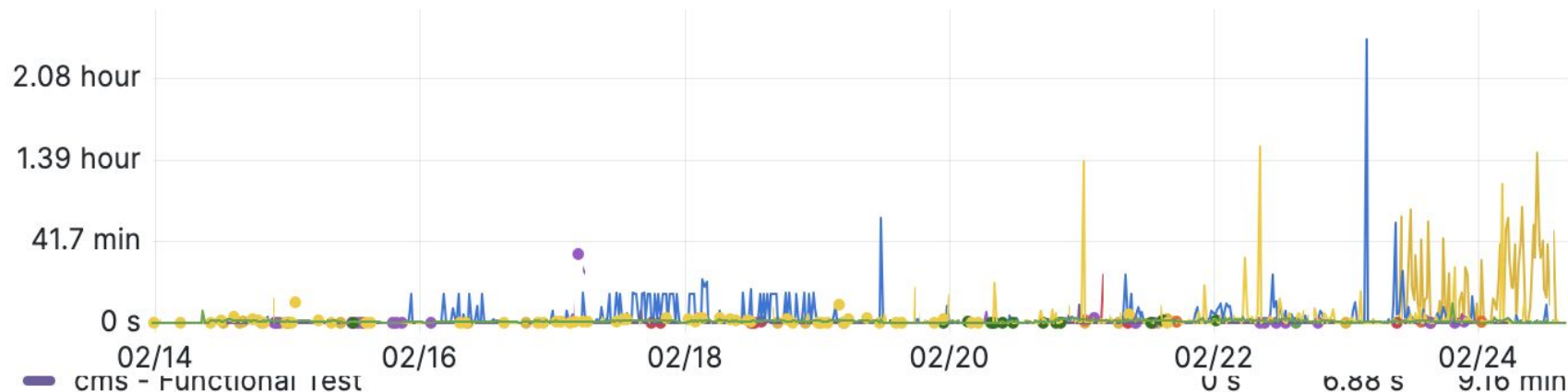


DC24: CERN EOS CMS

Median duration of flows split by Exp/Activity

Shows duration of DC flows was quite short wrt. production/rebalancing

Median Duration Received per Exp/Act



cms - Functional test	0 s	0.88 s	9.16 min
cms - Debug	0 s	11.6 s	2.08 min
cms - Data rebalancing	0 s	1.97 min	1.50 hour
cms - Data Challenge	0 s	46.2 s	9.93 min

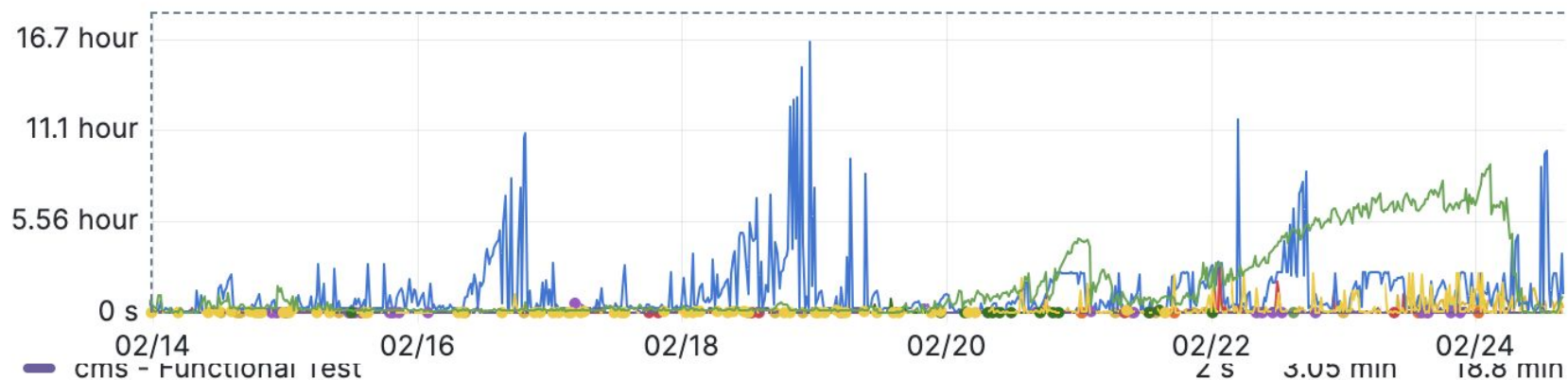
DC24: CERN EOS CMS

Max duration of flows split by Exp/Activity

First week had a lot of “fat” flows from production activity (but none from DC)

Second week was different, some DC flows took hours to finish

Max Duration Received per Exp/Act

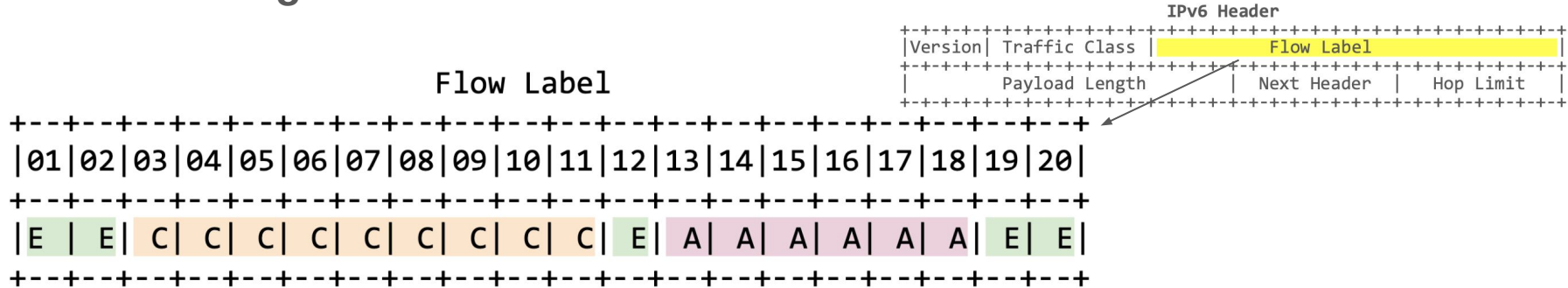


cms - Functional test	0 s	3.05 min	18.8 min
cms - Debug	0 s	2.93 min	52.5 min
cms - Data rebalancing	0 s	12.4 min	2.42 hour
cms - Data Challenge	1 s	1.59 hour	8.99 hour

Research & Development

Technical Spec for Packet Marking

Packet Marking via the use of the IPv6 Flow Label



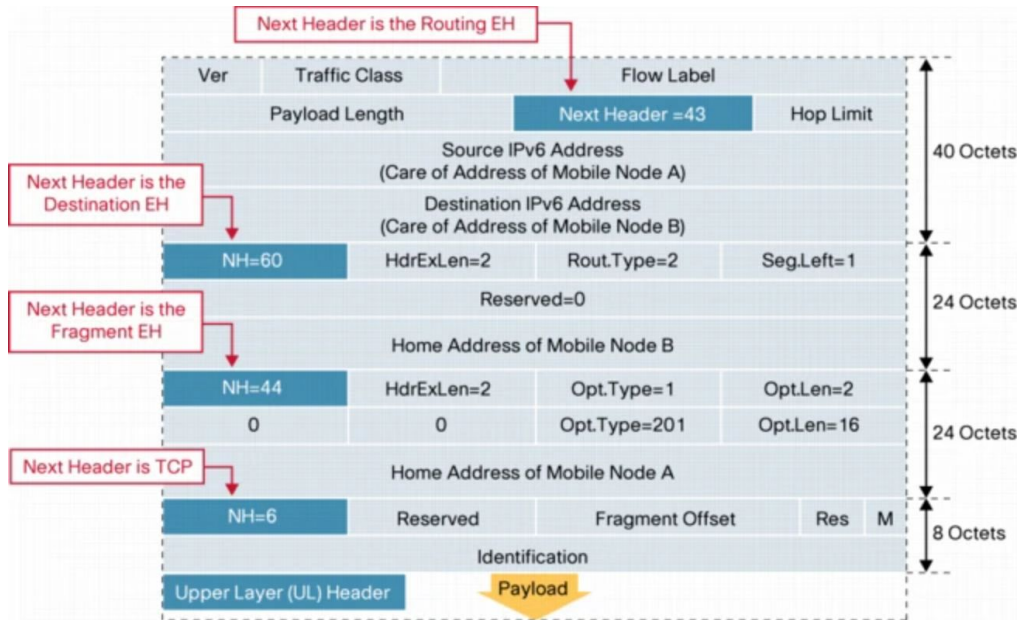
- (C) Community identifier: "Who are you affiliated with?"
- (A) Activity identifier: "What are you doing within your community?"
- (E) Entropy bits sprinkled throughout

[IETF RFC-Informational Draft](#) is available with more details

Destination Option and Hop-by-Hop Option

Based on feedback from IETF v6ops community

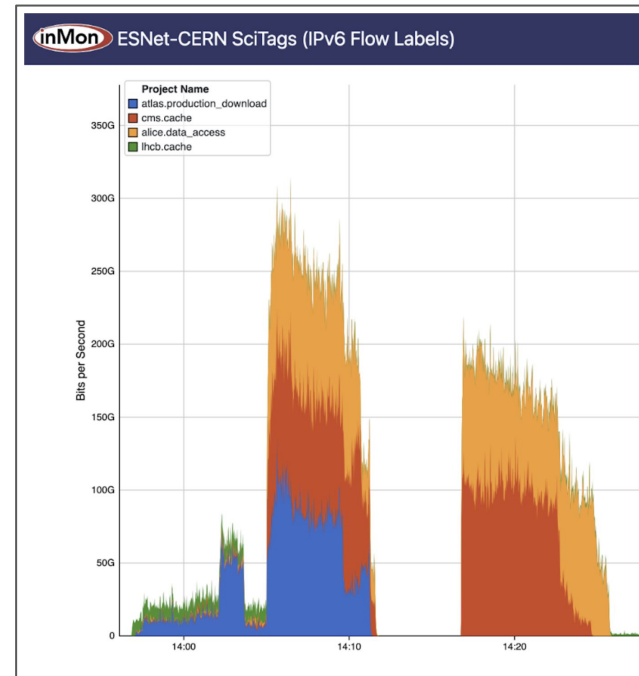
Started exploring Destination Option (DO) and Hop-by-Hop Option (HbH) as alternatives ([eBPF-PDM](#), [eBPF-extHeaders](#)) to flow label.



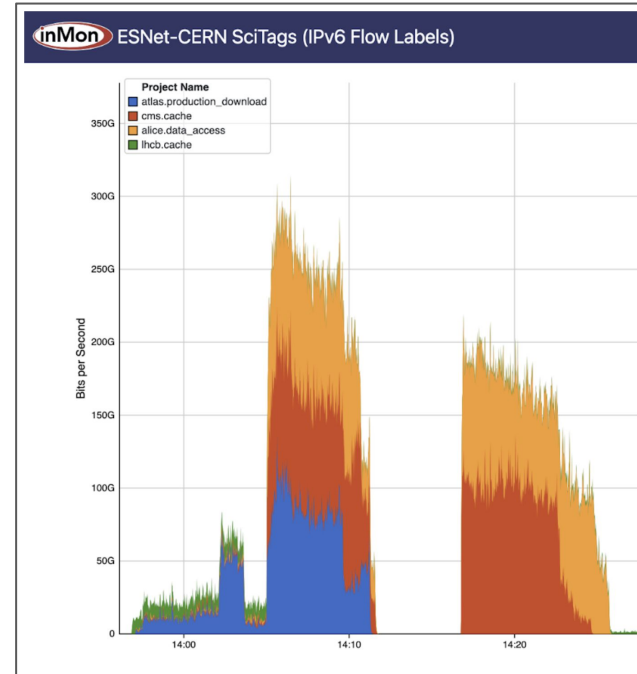
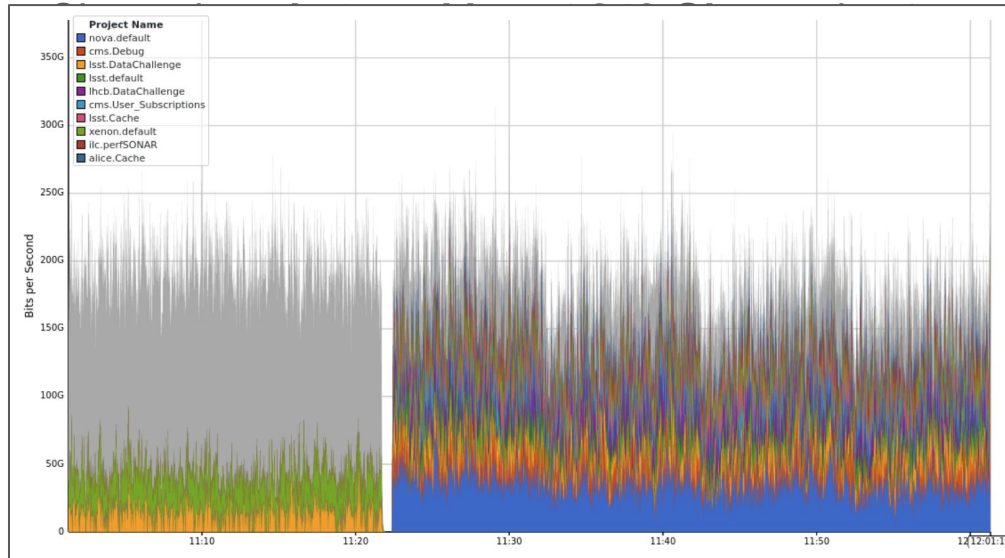
Packet marking demo with DO and HbH during SC24 Prototype implementation developed as part of **flowd** service

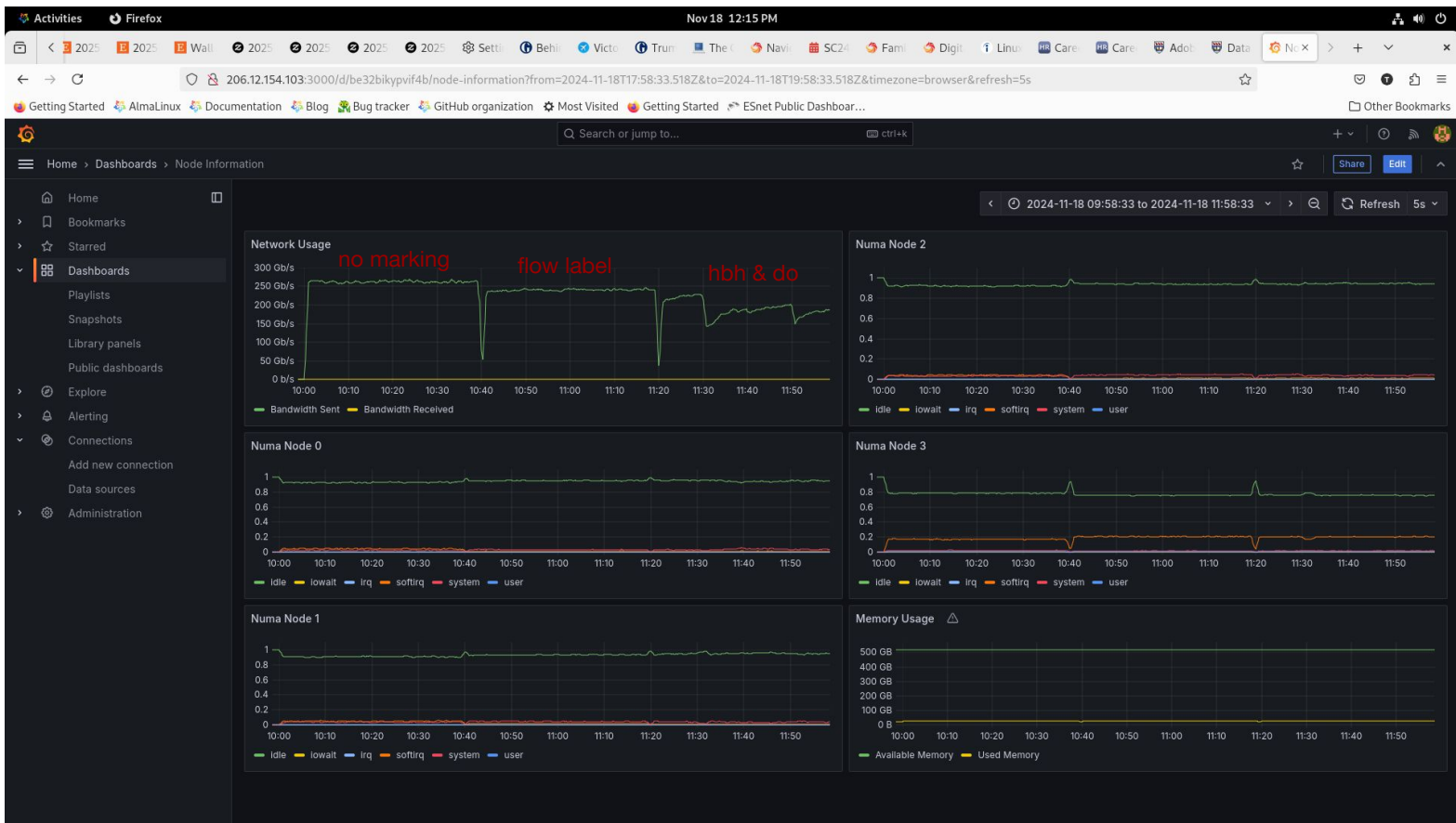
During Supercomputing 24 in Atlanta, we demonstrated a number of aspects of our packet and flow marking work.

- Showed **packet marking at 240 Gbps** using two GEN5 servers (with 400 Gbps NICs; peak rate without packet marking was 270 Gbps)
- In collaboration with inMon Corp, set up packet collectors [via sflow](#) and demonstrated **real-time monitoring of flows by community/activity**.
- Demonstrated packet marking using flow label as well as HbH and DO extension headers
- Deployment of flowd and packet marking on the [National Research Platform](#) (@UCSD)



During Supercomputing 24 in Atlanta, we demonstrated a number of aspects of our packet and flow marking work.



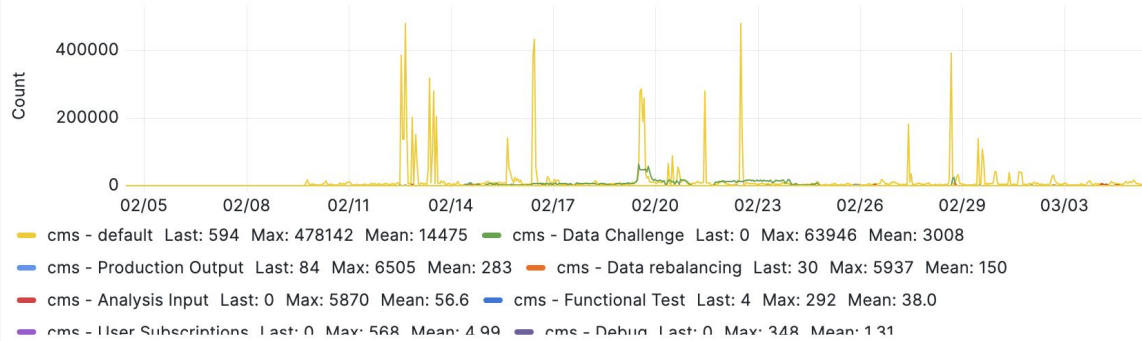


DC24: University of Nebraska (UNL) - many flows not coming from FTS

DC flows were only a small subset (IP split shows a very different profile)

Site UNL

Total Flows per Exp/Act



Total Flows per IP version

