GEANT P4 lab (aka GP4L)

Carmen MISA MOREIRA
CERN *– RARE project participant*
carmen.misa.moreira@cern.ch

TNC22

June 15th 2022

Public

www.geant.org

Agenda

GÉANT-4 Programme

GP4L Mission statement

A bit of context …

- Programming Protocol-independent Packet Processors: P4 language
- Network Operating System
- P4 Programmable Switches: EdgeCore Wedge100BF-32QS

GP4L - GÉANT P4 LAB footprint

GP4L - Use case at CERN

How can you benefit from GP4L ?

GÉANT-4 Programme

39 participants country

- NREN participants
- But also NREN staff members !

GN4 phase 3 cooperation programme

- 44 months
- Budget 78 M€

GN4-3 divided in “Work Package”

- 9 WP divided in “Tasks”
- Task 1 encompasses 6 Sub-tasks
- RARE/GP4L within a sub-task

GP4L Mission statement

- Primarily used to validate the software code inherently part of RARE/freeRtr open source routing stack.

- It is a service that is under development with the aim to provide experimental dataplane programming facilities where researchers can elaborate and test representative and geographically distributed network experiment
  - With the usage of RARE/freeRtr Operating System
  - Or simply use a clean slate environment.

    (i.e use exclusively **GP4L** without RARE/freeRtr dataplane & control plane)

A bit of context…

P4 language, Network Operating System and P4 Programmable Switches

A bit of context…
Programming Protocol-independent Packet Processors: P4 language

Language for programming the data plane of network devices

- Define how packets are processed
- P4 program structure: header types, parser/deparser, match-action tables, user-dened metadata and intrinsic metadata

Domain-specific language designed to be implementable on a large variety of targets

- Programmable network interface cards, FPGAs, software switches and hardware ASICs.

A bit of context…
Network Operating System

RARE/FreeRtr

- Controls the data plane by managing entries in routing tables
- Free and open source router operating system
- Export forwarding tables to DPDK or hardware switches
  - via OpenFlow or P4lang
- No global routing table
- Every routed interface must be in a virtual routing table

A bit of context…
P4 Programmable Switches

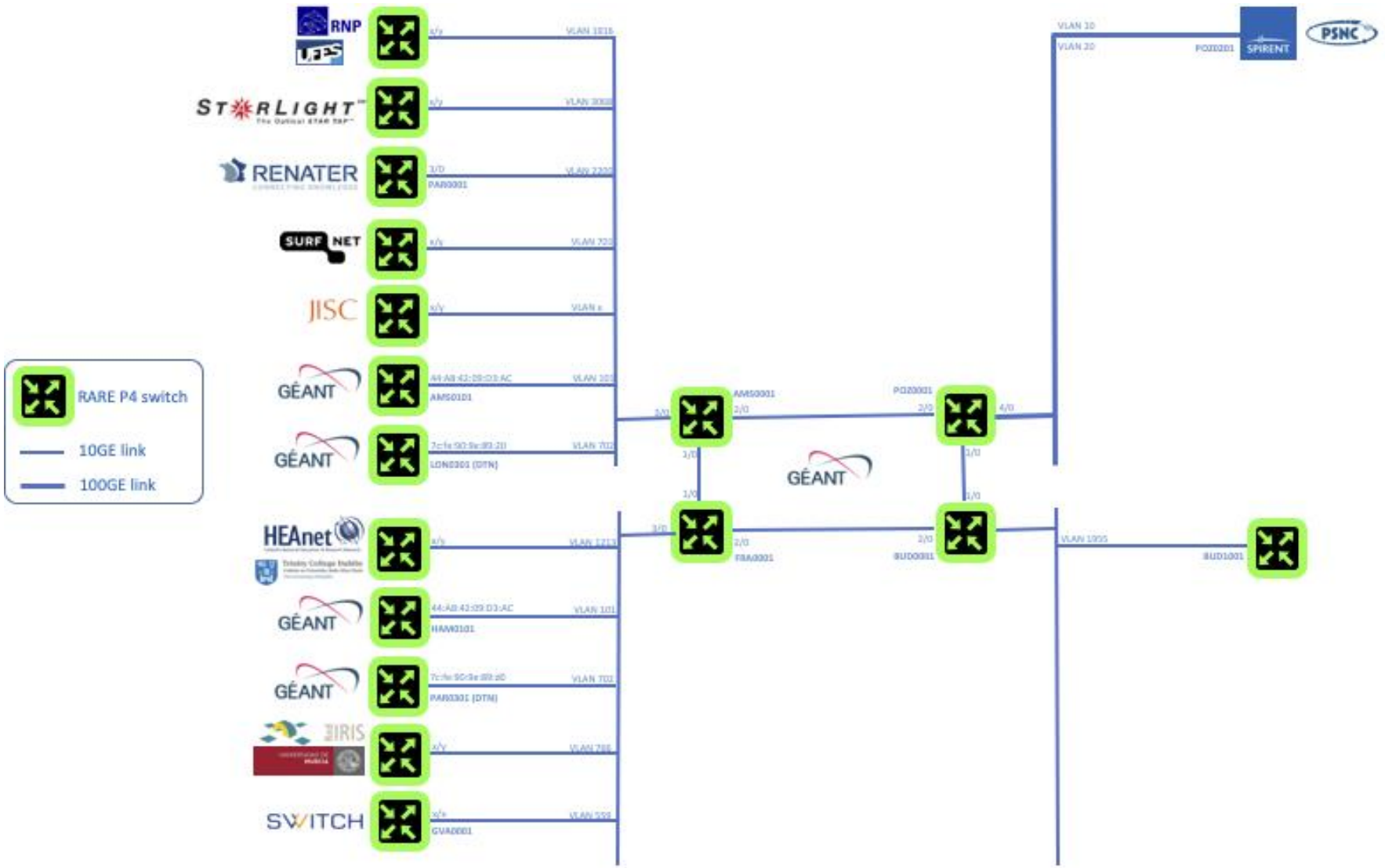
EdgeCore Wedge100BF-32QS:

- 100GbE Data Center Switch
  - Bare-Metal Hardware
  - L2/L3 Switching
  - 32xQSFP28 Ports
- Data-Plane Programmability
  - Intel Tofino Switch Silicon
  - Barefoot Networks
- Quad-Pipe Programmable Packet Processing Pipeline
  - 6.4 Tbps Total Bandwidth
- CPU: Intelx86 Xeon 2.0GHz
  - 8-core/48GB/2TB SSD

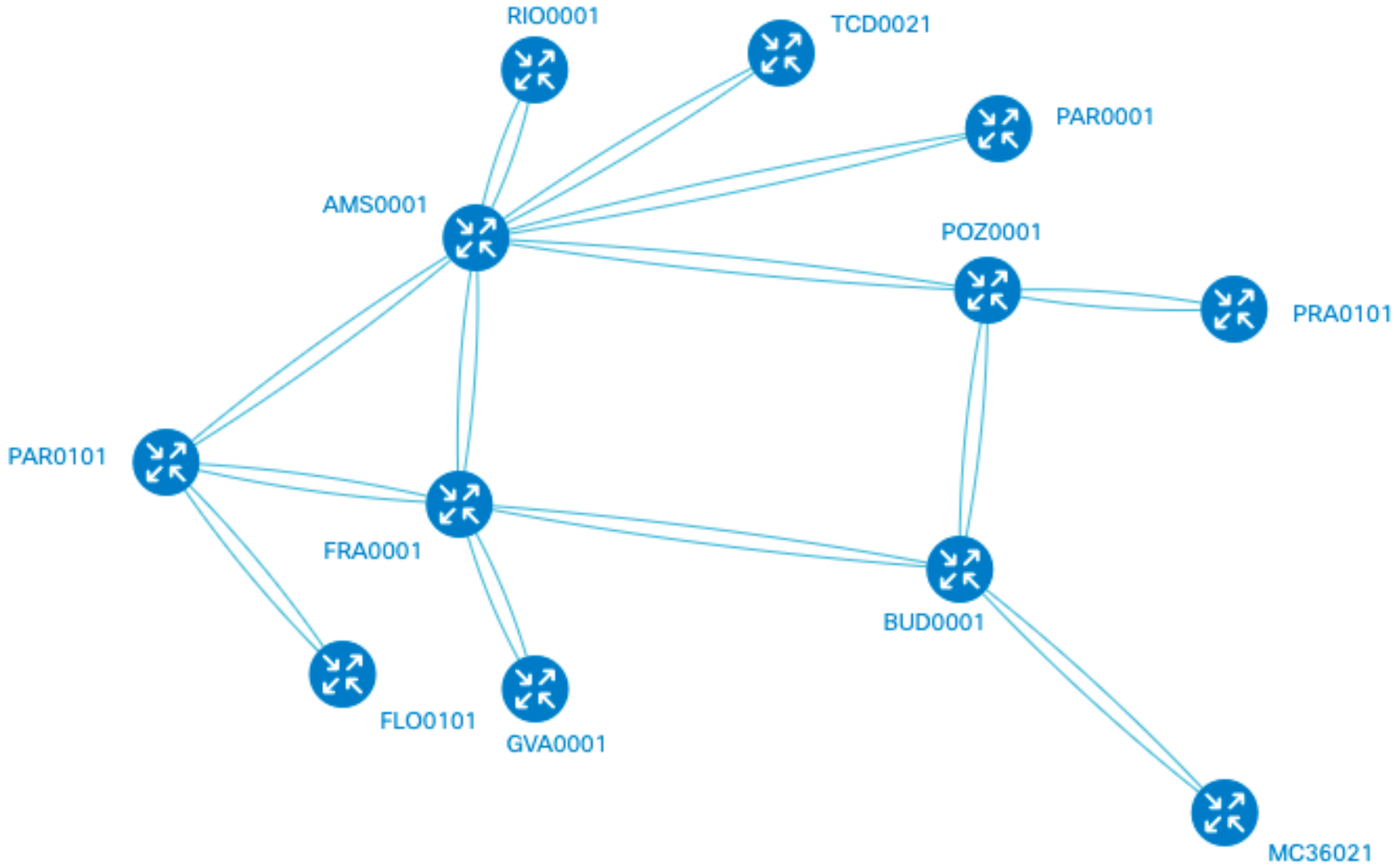
GP4L

GÉANT P4 LAB footprint

GP4L – GÉANT P4 LAB



GP4L – GÉANT P4 LAB: topology rendering via BGP-LS[1]



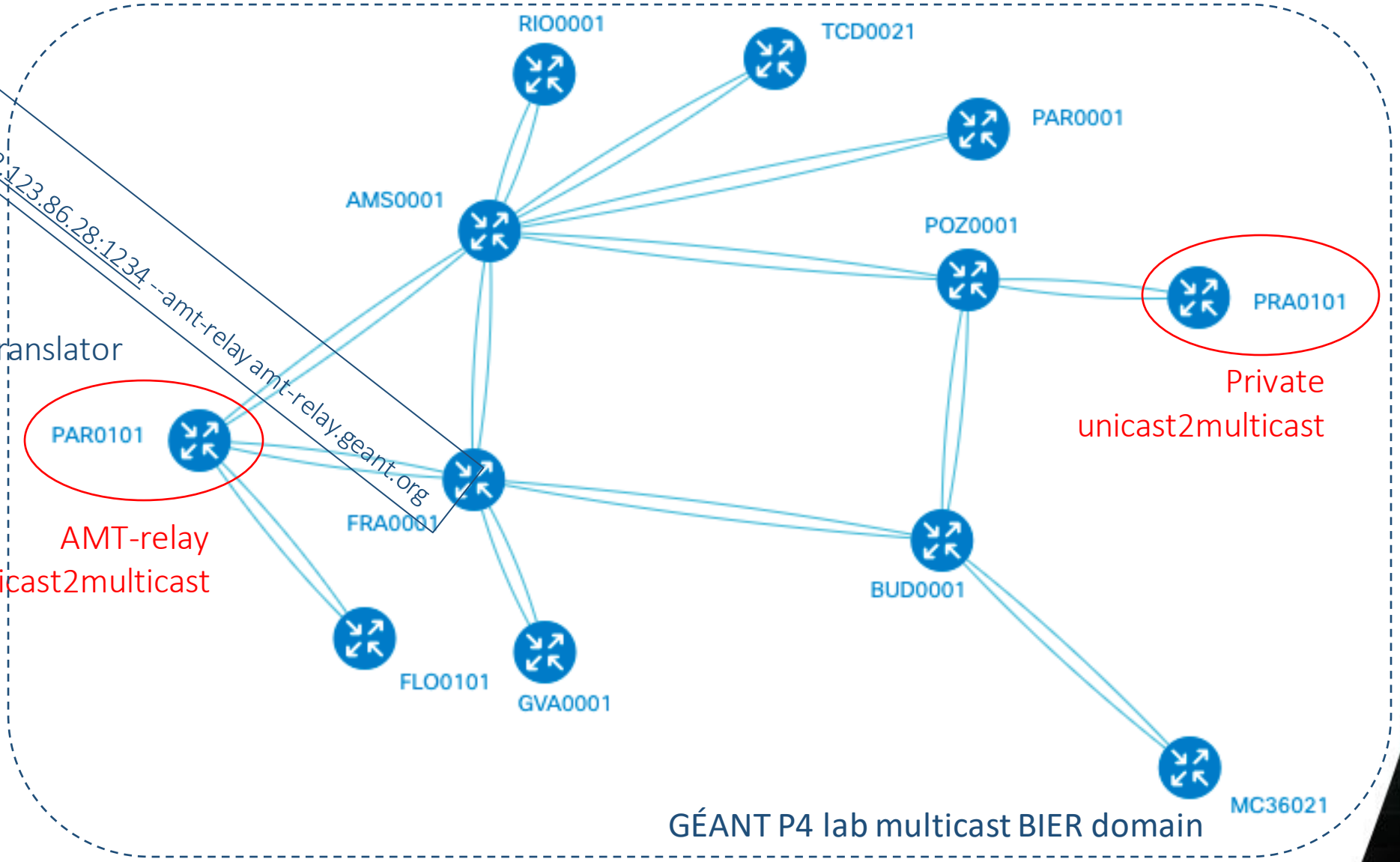
www.geant.org

[1] http://gp4l.geant.org/

GP4L AMT relay / AMT gateway / Unicast --> Multicast

[VLC 4 needed][2] --> vlc amt://62.40.109.31@232.123.86.28:1234 --amt-relay amt-relay.geant.org

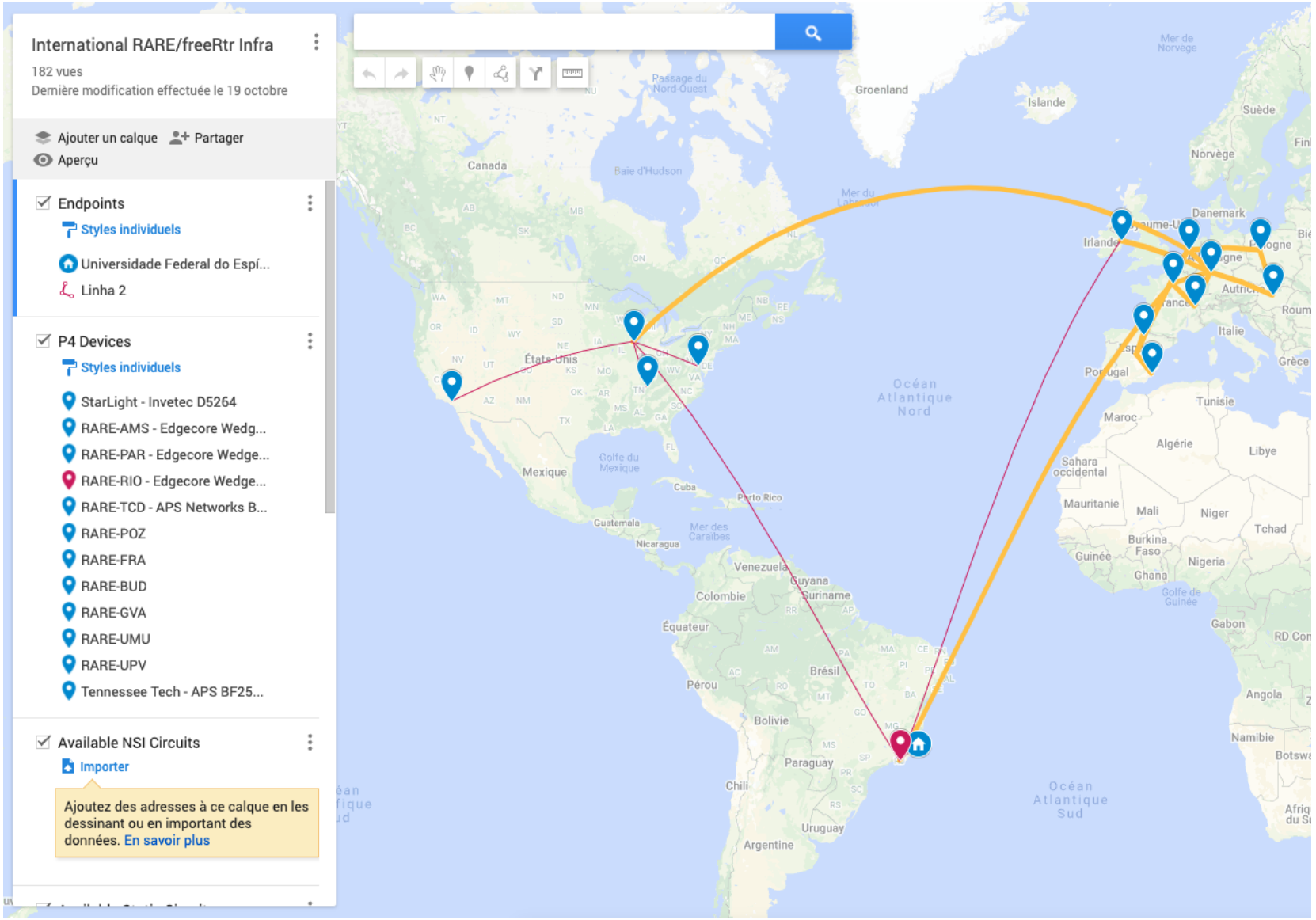
MULTICAST
BIER
AMT relay/gateway
Unicast2multicast translator



[1] http://gp4l.geant.org/
[2] https://nightlies.videolan.org/

AN INNOVATIVE MULTICAST STREAMING SERVICE !

GP4L – GÉANT P4 LAB



GP4L

Use case at CERN

Packet Marking Specification: IPv6 Flow Label

Research Network Technology WG proposed a packet marking technique to identify the LHC experiment and the application that has generated a transmission packet [1]

- The Experiment-Application tag is inserted in the flow label field of the IPv6 packet header.
- The original purpose of the field was never implemented, so it has been neglected by most of the existing network devices.

Primary goal of the marking is to count the traffic, but special routing polices could also be applied.

Flow label field of IPv6 header: 20 bits

- 5 entropy bits to match RFC 6436
- 9 bits to define the science domain
- 6 bits to define the application/type of traffic

Bits 12 - 13 Entropy	Bits 14 - 22 Science Domain	Bit 23 Entropy	Bits 24 - 29 Application	Bits 30 - 31 Entropy

```
Astro/HEP Science Domains:
Reserved - 0
Default  - 65536
ATLAS    - 32768
CMS      - 98304
LHCb     - 16384
ALICE    - 81920
Belle11  - 49152
SKA      - 114688
LSST     - 73728
DUNE     - 8192
```

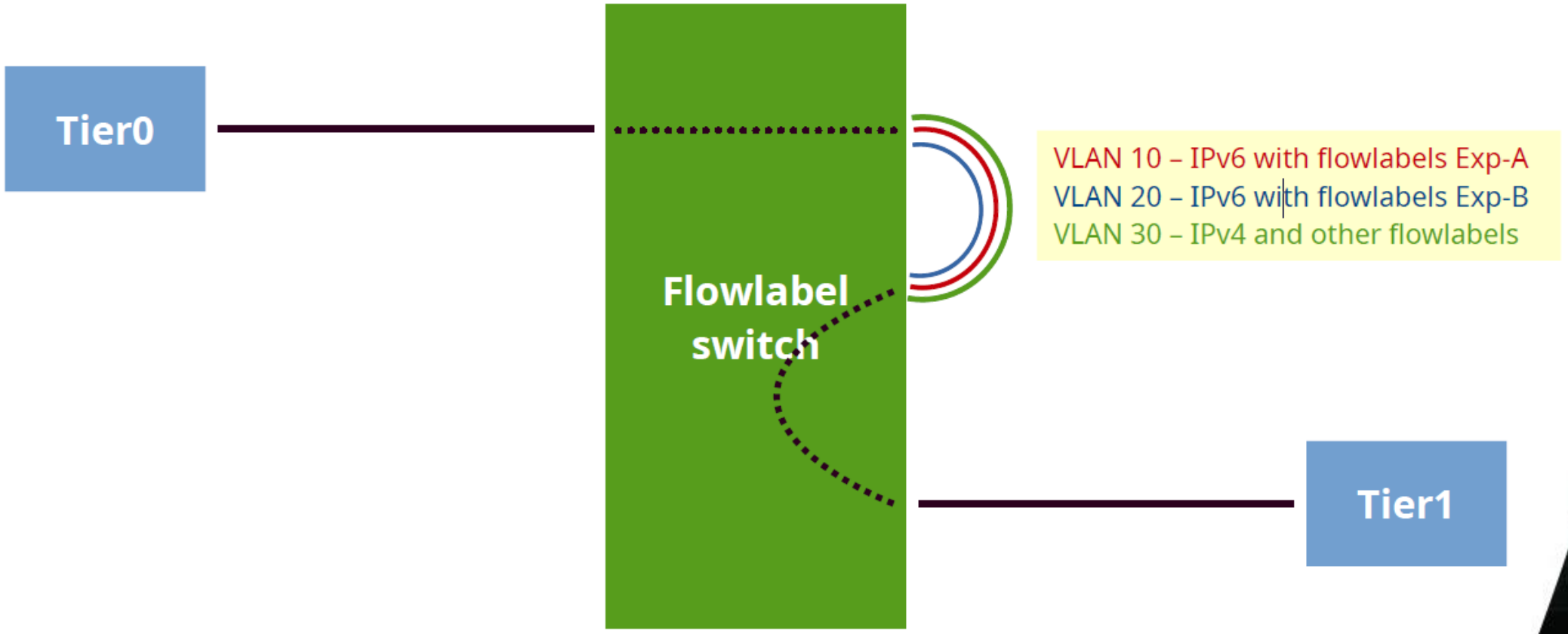
```
Application:
Reserved     - 0
Default      - 4
perfSONAR    - 8
Cache        - 12
DataChallege - 16
```

[1] www.scitags.org

Programmable Switches for Flow Label Accounting and Routing

Flow label is inserted transparently on LHCOPN link to a Tier1

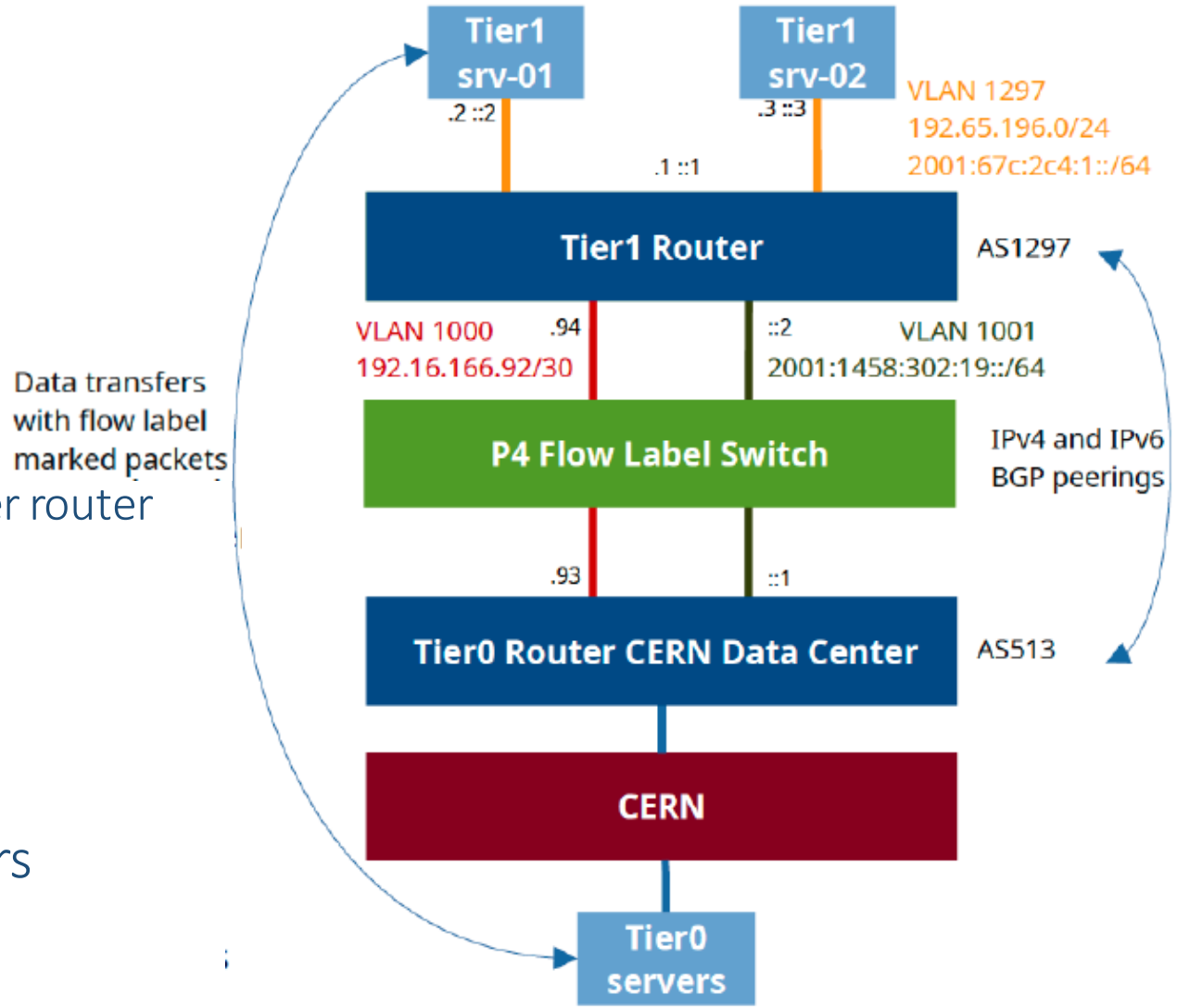
The traffic is separated-counted-aggregated internally



Programmable Switches for Flow Label Accounting and Routing

Network configuration:

- Emulates a Tier 1/0 link
- Tier1/0 routers
  - IPv4/IPv6 BGP peerings
- Tier0 router
  - LHCOPN production border router
- Pure layer 2 bridges
  - VLAN 1000: IPv4 traffic
  - VLAN 1001: IPv6 traffic
- Tier0 servers
- OpenStack product servers



Programmable Switches for Flow Label Accounting and Routing

P4 switch network configuration: pure layer 2 bridges

```
access-list acl_all_ipv6_flowlabels
     # Match <Experiment> and <ANY Application>
     sequence 10 permit all any all any all flow 131076 & 261884
     sequence 11 permit all any all any all flow 65540  & 261884
     sequence 12 permit all any all any all flow 196612 & 261884
     sequence 13 permit all any all any all flow 32772  & 261884
     # Match <Experiment> and <perfSONAR Application>
     sequence 20 permit all any all any all flow 131072 & 261632
     sequence 21 permit all any all any all flow 65536  & 261632
     sequence 22 permit all any all any all flow 196608 & 261632
     sequence 23 permit all any all any all flow 32768  & 261632
     # Permit the rest of the traffic
     sequence 30 permit all any all any all
     exit
interface sdn1.1000
     description [VLAN ID=1000]
     bridge-group 1
     no shutdown
     no log-link-change
     exit

interface sdn1.1001
     description [VLAN ID=1001]
     bridge-group 2
     bridge-filter ipv6in acl_all_ipv6_flowlabels
     no shutdown
     no log-link-change
     exit
```

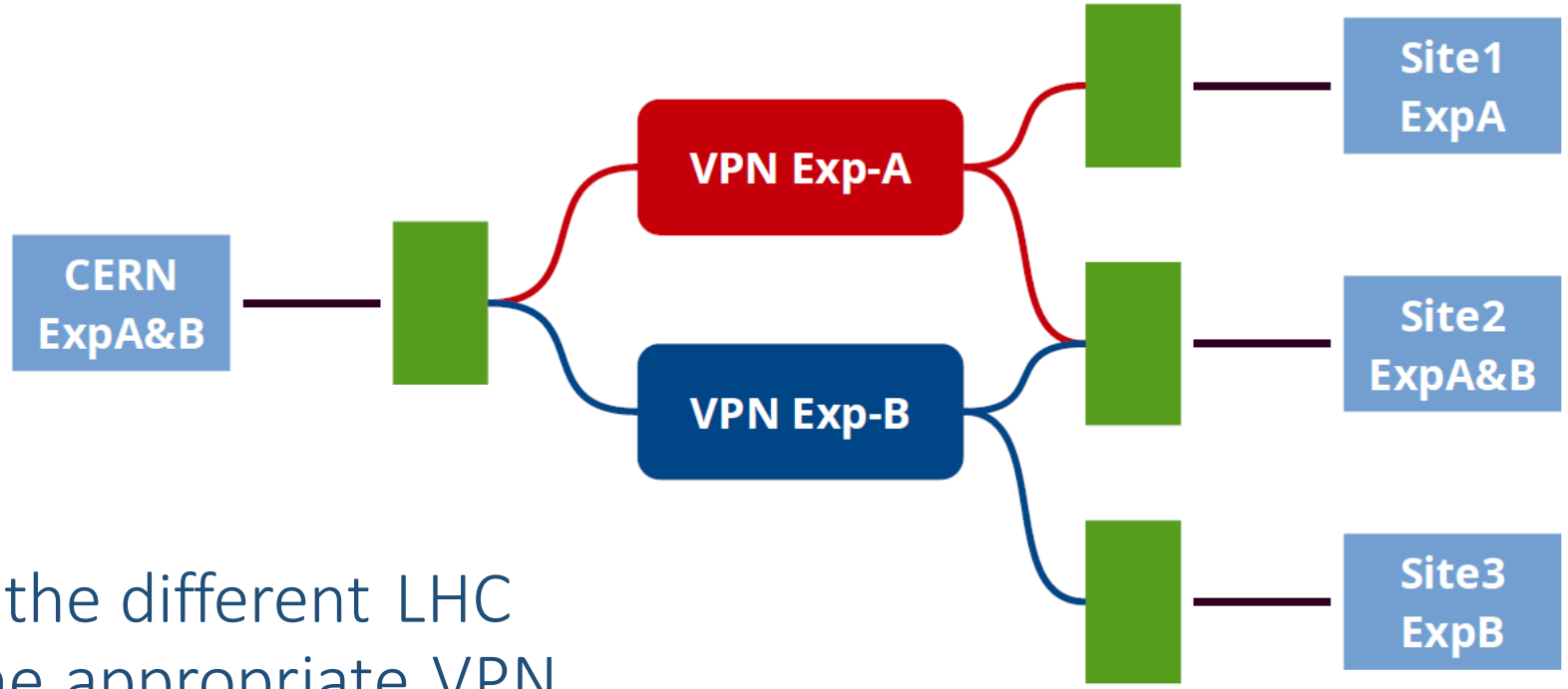
ATLAS <any>
CMS <any>
LHCb <any>
ALICE <any>

ATLAS <perfSONAR>
CMS <perfSONAR>
LHCb <perfSONAR>
ALICE <perfSONAR>

VLAN 1000 belongs to bridge 1

VLAN 1001 belongs to bridge 2
Filter IPv6 traffic at the input based on the access-list sentences

MultiONE multiple "LHCONEs": Traffic separation with IPv6 flow labels



Route the traffic of the different LHC experiments into the appropriate VPN.

- A prototype of a flow label router is being developed using a P4 programmable switch (EdgeCore Wedge100BF-32QS with Intel Tofino processor)

Useful links

Project:

GP4L project page: https://wiki.geant.org/display/GP4L/Home

RARE/freeRtr: https://www.rare.freertr.net

https://docs.freertr.net

https://blog.freertr.net

Contact:

Users: gp4l-users@lists.geant.org

Developer wanabee: gp4l-dev@lists.geant.org

Project: gp4l@lists.geant.org

Looking ahead: Finalize transition to production

Validate your use case with GP4L !

Orchestrate and automate GP4L:

- Lab reservation
- Persistent testbed interaction at global scale

New hardware:

- TOFINO2
- NVIDIA DPU
- P4 SmartNIC
- TOFINO/FPGA

Global worldwide footprint:

- Interconnection with other persistent testbed

➔ New idea:

- Validate new use cases
- Scalability
- 100/400 GE DTN automation
- Control plane scalability

And more …

Thank you

Any questions?

www.geant.org

The research leading to these results has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No. 856726 (GN4-3).