RNP's experiences with P4

**Marcos Schwarz – R&D Manager in Ciberinfrastructure at RNP
GNA-G AutoGOLE / SENSE WG Co-Chain**

March 30-31th – GNA-G Community VCs Q1 2022

Agenda

OpenRAN / ODISSEIA project (Testbed based on ONF Aether/SD-FABRIC)

P4 Operations (GNA-G / RARE Cooperation)

Multi NOS image

Integrating protocols defined by R&E community: RARE/freeRtr + PolKA

OpenRAN / ODISSEIA Project

- R&D on technologies that enable softwarization of network infrastructures in multi-domain (WDM, FTTH/PON, Packet/P4, 5G/OpenRAN, Edge Cloud) using open and disaggregated solutions

- Based on ONF's Aether project
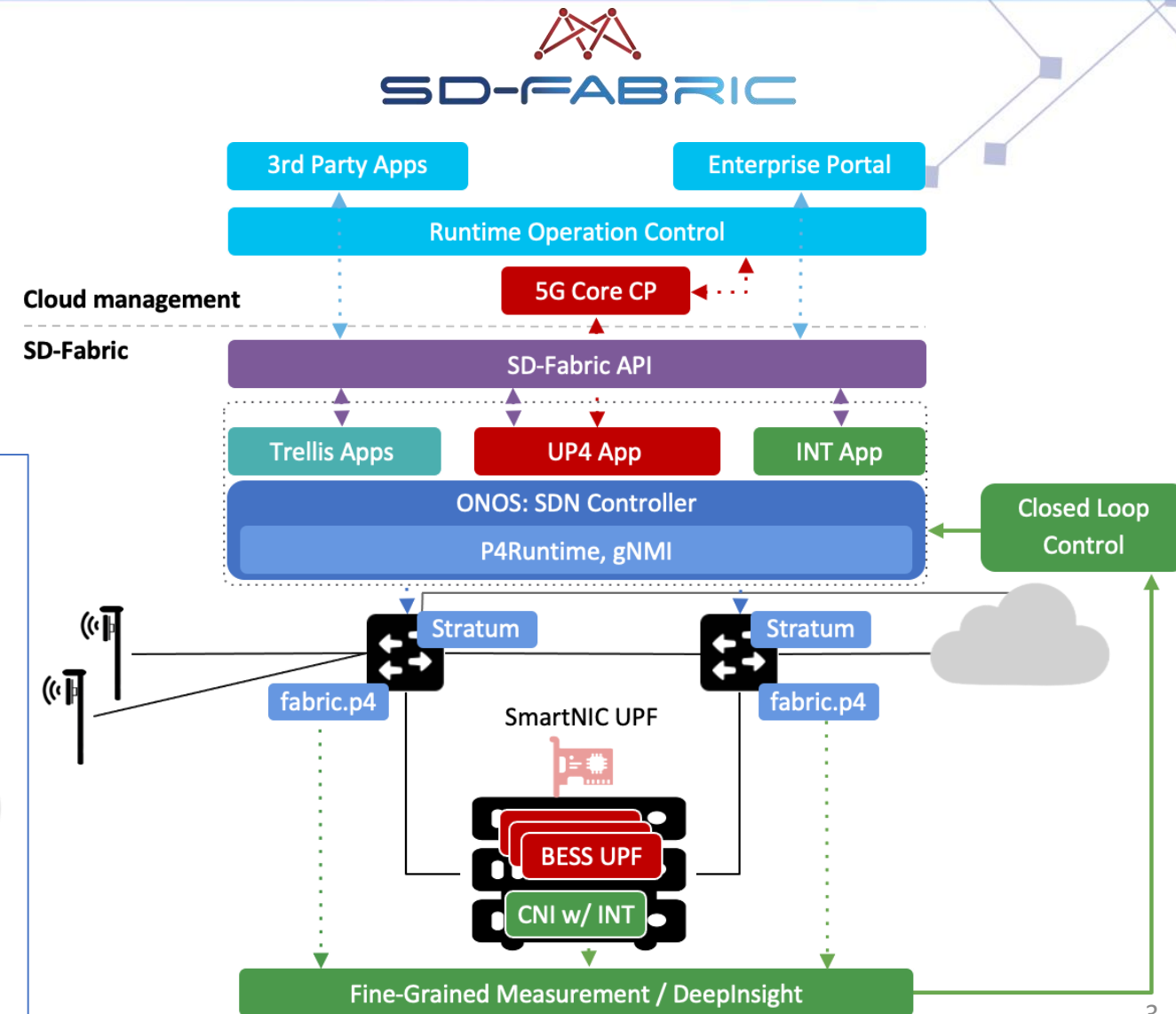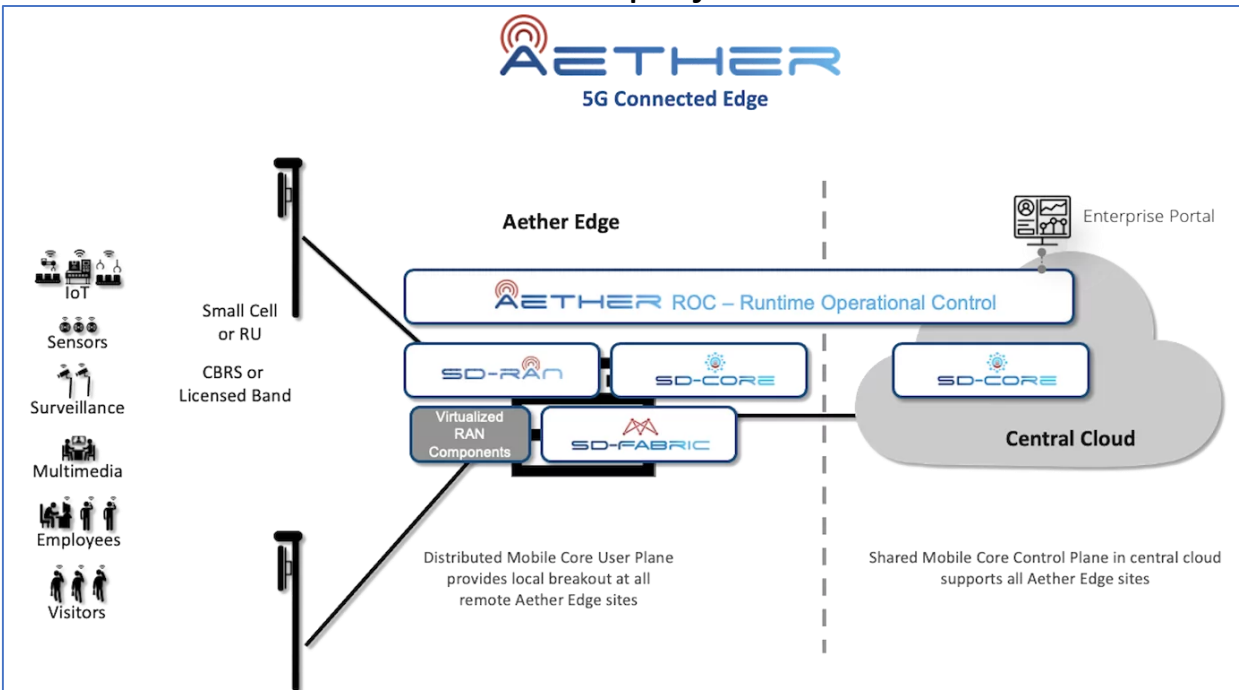
Persisten Multi-Resource Infra

Pre-production programmable network
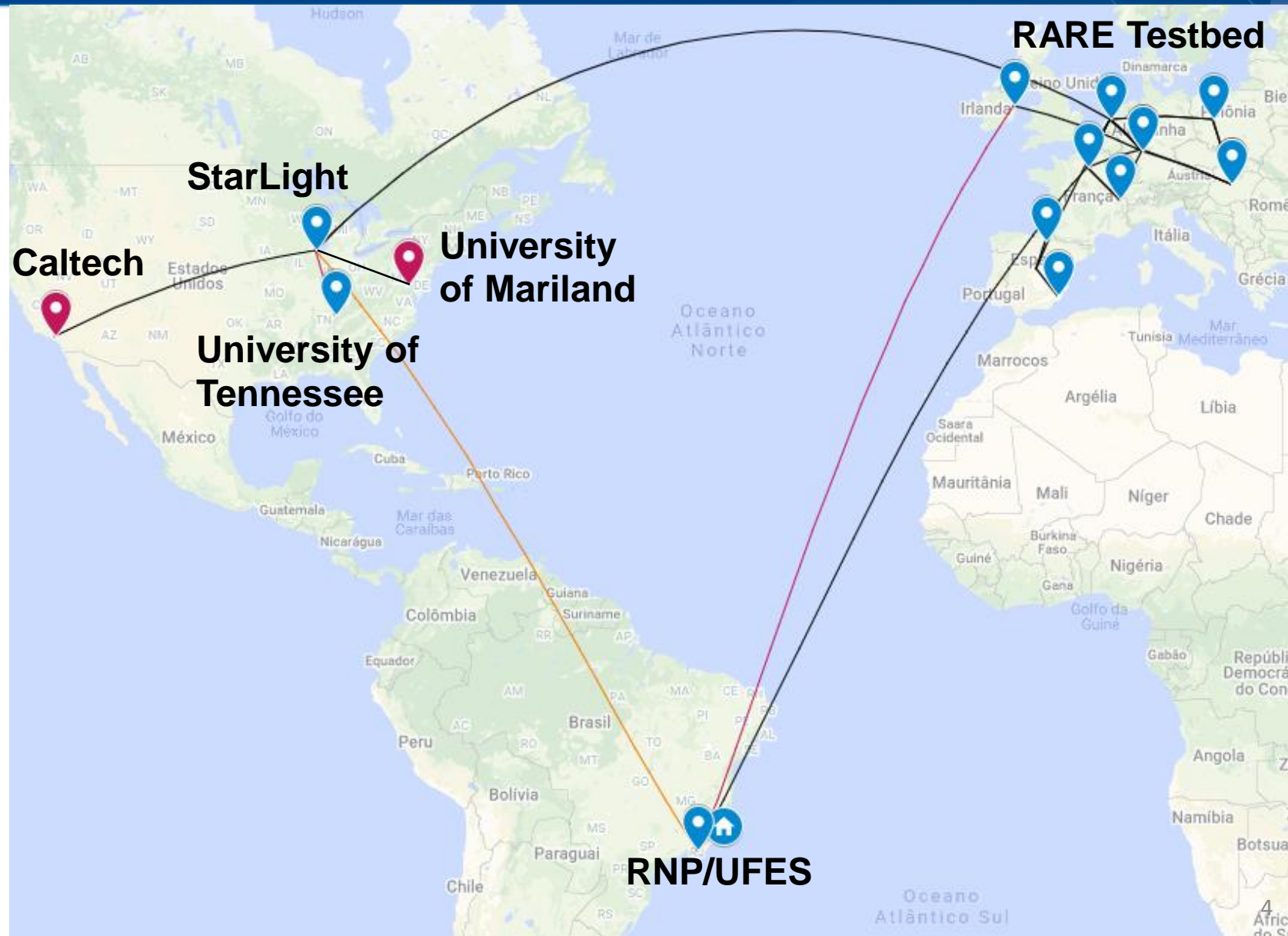to composse/validate new services

- Packet Marking
- INT based Congestion Detection
- Flow Steering

Joint Initiative

- GNA-G Data Intensive Science
- GNA-G AutoGOLE / SENSE
- RARE project

Resources / Services

- 100G DTNs and P4 Switches
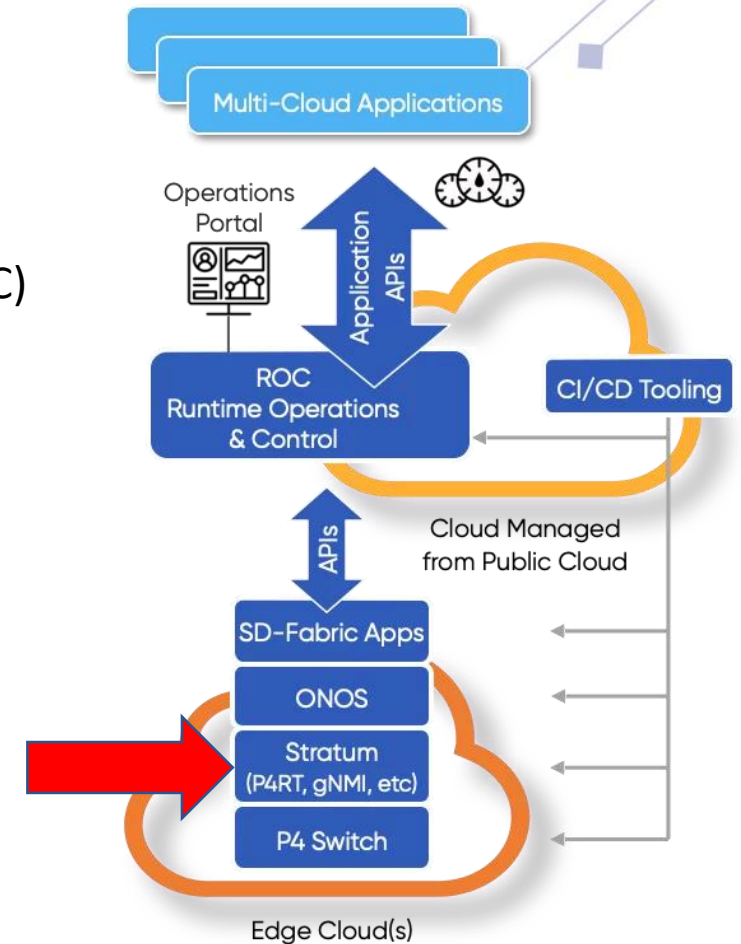- L2 Circuits and DTN orchestration

Multi NOS Image

- Provide an integrated image to facilitate experimention on multiple P4 Network Operating Systems

- V1 based on PINS (SONiC) image, enhanced with Stratum and RARE/freeRtr

- Initially supported platform EdgeCore Wedge100G (Tofino)

- Documentation to be shared among GNA-G AutoGOLE/SENSE members, on how to build the image and basic examples on how to use each NOS

- Future plan include exploring Tofino multi pipelines to run multiple Oses simultaneosly

Open source, production targeted, thin switch OS

- There's no routing (BGP), trunking (LACP), etc. by default
- Pairs with fabric.p4, ONOS, and Trellis for a complete networking stack (SD-FABRIC)

Next-Generation SDN
Complete set of next-generation interfaces for comprehensive design & runtime instrumentation

Pipeline 'Contract' Definition
Expressed in P4 Language (or possibly another language)

On compatible systems can be updated at runtime

Northbound

Embedded System

Pipeline Control

P4Runtime

Configuration

gNMI with OpenConfig

Operations

gNOI

Packets

programmable | pipeline

match | action | steps

Multi-Cloud Applications

Operations Portal

Application APIs

ROC Runtime Operations & Control

CI/CD Tooling

APIs

Cloud Managed from Public Cloud

SD-Fabric Apps

ONOS

Stratum (P4RT, gNMI, etc)

P4 Switch

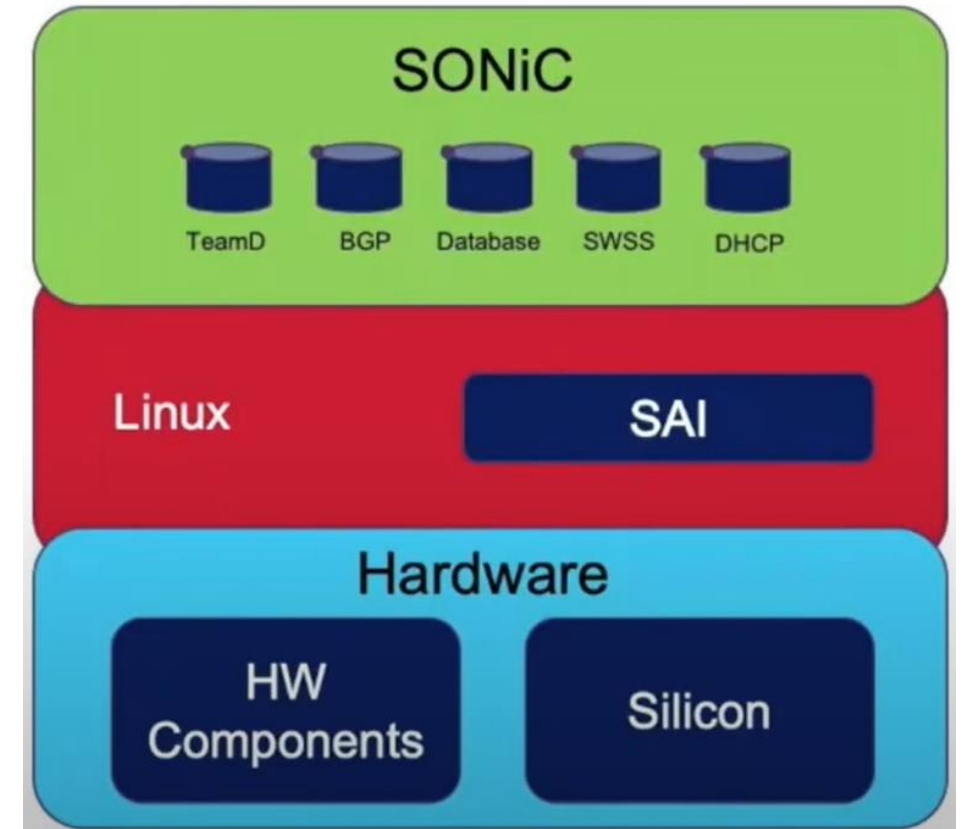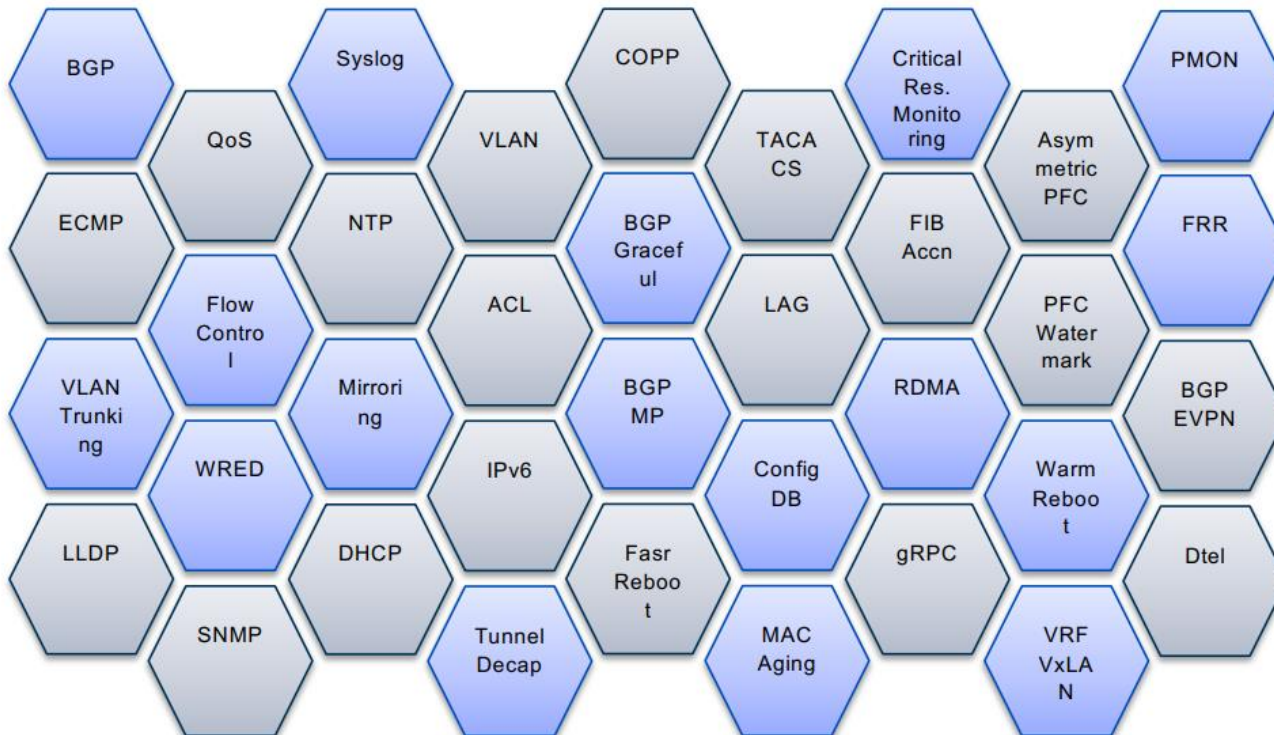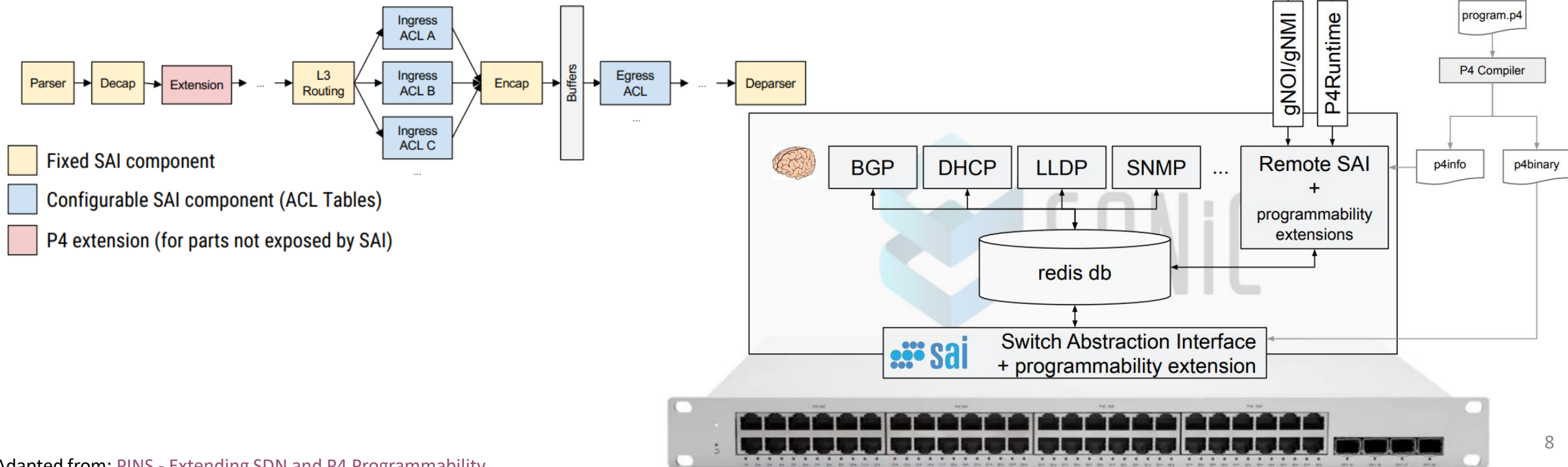Edge Cloud(s)

6

- Open and multi-vendor Network Operating System focused on datacenter requirements/features
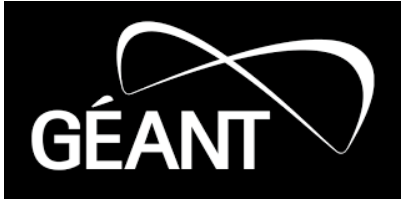- Support for traditional (fixed-function) and programmable switches

Hybrid control plane with path to SAI extensibility (SONiC + Stratum)

- Great for traditional network users that want some SDN capabilities
- Seamless transition for SONiC users

Fixed SAI component

Configurable SAI component (ACL Tables)

P4 extension (for parts not exposed by SAI)

RARE (Router for Academia, Research & Education) focus on creating an Open Source routing software platform.
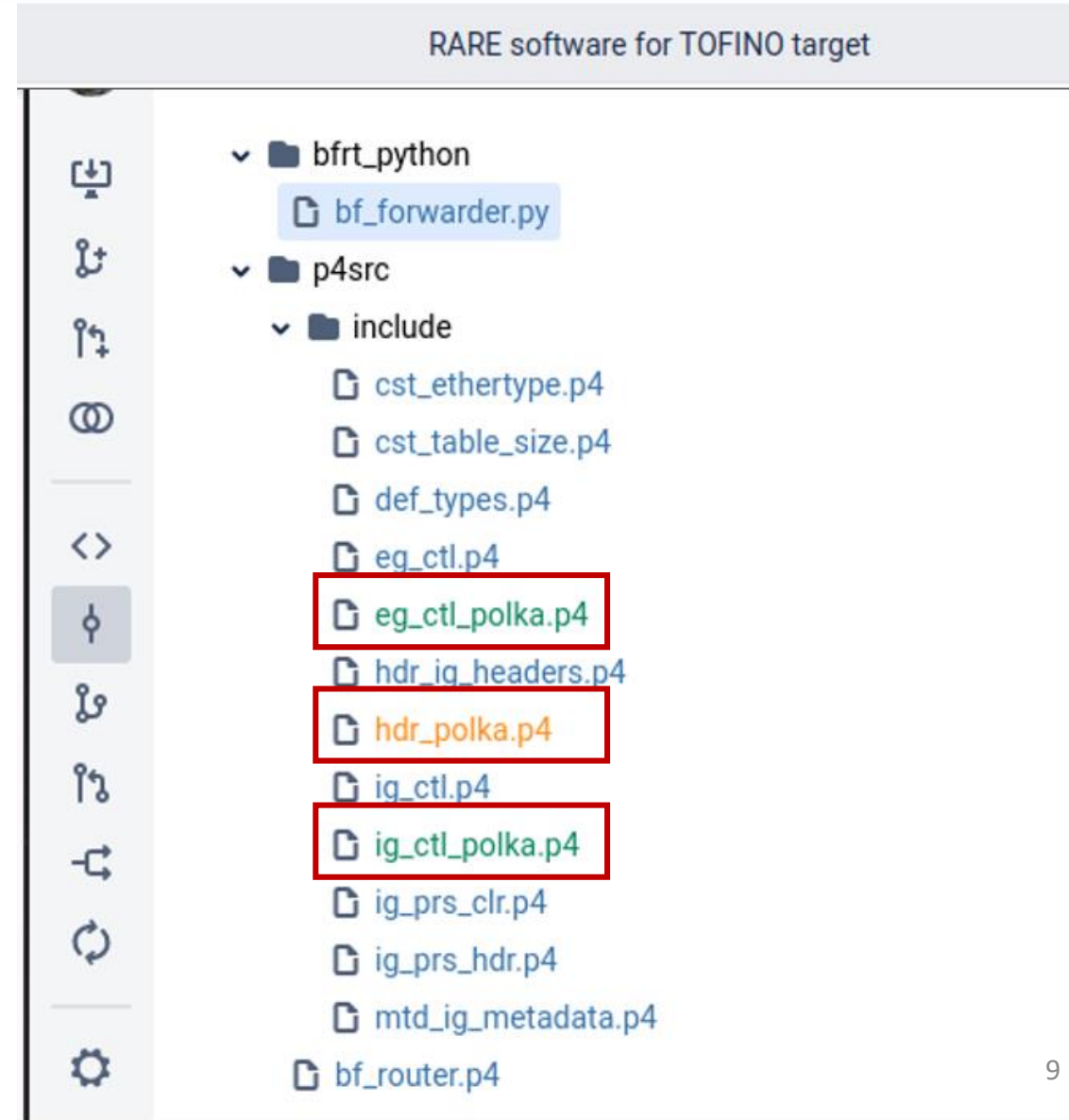
Multiple backends
- Tofino, DPDK, eBPF

Focus on R&E community requirements
- Router, BNG, SOHO
- Source-based routing (PolKA), Packet Marking

Modularization
- Unit tests for each protocol/feature
- Simple pipeline for integration/extension
- Enables parallel development

RARE software for TOFINO target

- bfrt_python
  - bf_forwarder.py
- p4src
  - include
    - cst_ethertype.p4
    - cst_table_size.p4
    - def_types.p4
    - eg_ctl.p4
    - eg_ctl_polka.p4
    - hdr_ig_headers.p4
    - hdr_polka.p4
    - ig_ctl.p4
    - ig_ctl_polka.p4
    - ig_prs_clr.p4
    - ig_prs_hdr.p4
    - mtd_ig_metadata.p4
  - bf_router.p4

PolKA: Polynomial Key-based Architecture

- A **Source Routing** approach that simultaneously meets the requirements:

| topology agnostic | no tables in the core | implementable in prog. switches (CRC hardware) | encoded path |
| | multipath expressiveness | | fixed header |
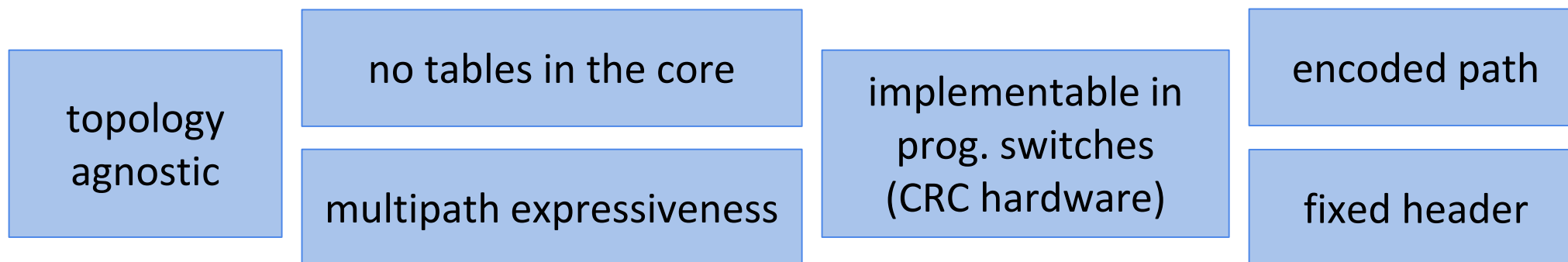
- Polynomial Residue Number System (RNS)

- Forwarding based on remainder of division:

$$portID = < routeID >_{nodeID}$$

$1 \quad = \quad <10000>_{0011}$

$10 \quad = \quad <10000>_{0111}$

$110 \quad = \quad <10000>_{1011}$

Source: PolKA: Polynomial Key-based Architecture

Timeline

2020	2021/1	2021/2	2022	
PolKA paper IEEE NetSoft	**ONDM paper Deploy @RARE**	**Integration with RARE+FreeRouter**	**M-PolKA paper IEEE TNSM** *(early access)*	**PolKA@pangr IETF 113**
Novel **Polynomial RNS-based SR and reuse of CRC hardware**		PolKA **data & control plane implementation** + **integration**	Extension to **multipath SR** for reliable communications	**Lightning Talk**
Emulated prototype in Mininet	Hardware prototype in Intel Tofino	Emulated prototype in FreeRouter & Hardware prototype in Intel Tofino with FreeRouter control plane	Innovative apps: inband network telemetry, and load balance	Next steps: produce and submit IETF draft

Technical Presentation on April 14th at GNA-G Data Intensive Sciences WG meeting (14:00 GMT)

Thanks! Questions?

Marcos Schwarz

marcos.schwarz@rnp.br