SC22 Network Research Exhibition: Demonstration Preliminary Abstract

Demonstrating PolKA routing approach to support traffic engineering for data-intensive science

Everson Scherrer Borges, Cristina Dominicini, Rafael Guimarães,

Edgar Pontes, Magnos Martinello and Moisés Ribeiro

IFES {everson,cristina.dominicini,rafaelg}@ifes.edu.br

UFES {edgar.pontes,magnos.martinello, moises.ribeiro}@ufes.br

Abstract

The current requirements of globally distributed workflows of the LHC and other data-intensive science programs, and the challenging projections for the next years, indicate the urgent need for new approaches to address the balance between innovative functionality, performance, cost, and other policy-driven factors when it comes to data transport across networks. The High Luminosity LHC (HL-LHC) experiments alone, for example, require terabit/sec data flows across regional, national, and intercontinental network paths by around 2028. In addition, the Vera Rubin Observatory, DUNE at LBNF, the Square Kilometer Array Telescope, and others will bring not only extra traffic volumes to be handled across such paths, but also new Traffic Engineering (TE) challenges in order to manage and control such a diverse set of requirements of coexisting and competing for flows across a complex intercontinental network topology.

This NRE proposes to demonstrate PolKA functionalities to support the TE extreme challenges for data-intensive science. PolKA is a novel source routing approach that explores the Residue Number System (RNS) and Chinese Remainder Theorem (CRT) by performing the forwarding as an arithmetic operation: the remainder of division. PolKA encodes the path in a routeID using the RNS in contrast to the conventional list-based representation, which transports the path information “in clear” inside the packet header. Then, PolKA core nodes use this encoded route label to discover the output ports.

The plan of demonstration is to create an overlay network with PolKA tunnels forming virtual circuits to validate the data-intensive transfer over 10G and 100G+ as a proof-of-principle of PolKA mechanisms. At the edge, flows can be classified, balanced and steered by using a Policy-Based Routing (PBR). A number of virtual circuits may be configured by dividing the capacity of the physical links and using them to serve the flows. Underlay congestion can be detected by tunnel monitoring and signalized to the overlay, and overlay routing can steer traffic from congested tunnels to other paths. Comparisons between segment routing and PolKA regarding controllability and performance metrics are also planned in this proposal.

RARE (the Router for Academia, Research & Education) is a GÉANT project which is developing and deploying an open-source routing software platform solution that provides innovative functionality, interoperability, and high-performance data planes. Network solutions are sought out from the perspective of National Research and Education Networks (NRENs), and not from regular service operators', which makes RARE's outcomes very relevant to GNA-G's goals. RARE subscribes to the new network paradigm of software-defined networking, which is based on a separation between the control plane and data plane to address the innovative functionality issue. Along with a Free and Open Source Software (FOSS) system on the control plane called FreeRouter(freeRtr), RARE manages to re-conciliate open innovation with interoperability with legacy protocols and networks, and adapt to and interwork with key technologies emerging in the network industry. For performance, RARE/freeRtr relies on the P4 open-source network programming language in order to expedite packet processing exploiting cutting-edge programmable ASICs embedded in modern network boxes. It also supports the P4Runtime API, which is a control plane specification for controlling the data plane elements. Last but not least, FOSS will address the pressing cost issues for acquiring fully-featured operational systems versions of high-end network equipment.

Goals

The goal of this proposal is to investigate whether PolKA approach deployed at RARE/freeRtr meets the needs of DIS networks, working with other software tools and subsystems developed by the DIS-WG for constructing switched overlay networks composed of network paths with bandwidth guarantees, load balancing, prioritizing and scheduling flows over selected multi-domain paths, and making decisions on the coordinated use of network and site computing and storage resources to help accelerate the science workflows.

In particular, this NRE will focus on providing extreme TE functionalities allowed by FOSS versions of emerging industry-driven Source Routing Techniques (SRT), such as Segment Routing, and compare those with cutting-edge academic developments. In the latter category, **Pol**ynomial **K**ey-based **A**rchitecture (**PolKA**) for Source Routing in Network Fabrics was chosen since it might be a better enabler for orchestrating computing and storage resources globally among DIS facilities. Our purpose is to:

1. Create a PolKA overlay by integrating existing resources from GNA-G AutoGOLE/SENSE testbeds that allows simultaneous execution and comparison with Segment Routing.
2. Design and execution of emulated experiments based on GNA-G topologies with freeRtr environment to compare PolKA and Segment Routing functionalities.
3. Implement the emulated topologies and traffic scenarios on the RARE/freeRtr testbed with the Barefoot Switch in order to select a representative proof-of-principle case.
4. Validate a data-intensive transfer over 10G and 100G+ as a proof-of-principle of PolKA mechanism on the GNA-G AutoGOLE/SENSE persistent testbed to avoid congestion and overload of international links capacity.

Controllability enabled by source routing

The most traditional way of executing source routing is to represent the path as a list of output ports and the forwarding operation as a pop. Although the most disseminated source routing protocol is Segment Routing, it has some limitations: i) it depends on expensive equipment and proprietary implementations; ii) its MPLS version still depends on tables in the core nodes; iii) it depends on variable-length headers, which limits the number of maximum hops implementable in hardware switches; and iv) there is no direct multicast support.

PolKA is a novel source routing approach that explores the RNS and CRT by performing the forwarding as an arithmetic operation: the remainder of division. PolKA can be deployed in high-performance P4-enabled programmable switches with the reuse of the CRC hardware, and its performance is equivalent to traditional approaches. Thus, this special path encoding and routing mechanism allows PolKA to offer the following advantages: i) it does not keep any table in the core network; ii) the packet header has fixed length and the route label does not change throughout the path; iii) it can represent multipaths in the network layer for any topology.

Resources

RARE/freeRtr has a wide range of legacy protocol FOSS components and subsystems with (operator-class) validated implementations. It can be used in all-in-one installations in order to emulate large networks for testing and validating new functionalities. Moreover, it has a unique seamless portability process (i.e., from emulation setups into testbed experiments) for performance test validation on different physical data planes. Finally, it should be highlighted that freeRtr is one of the few FOSS router operating systems that have segment routing implemented; and has a full implementation of PolKA. Thus, RARE/freeRtr suits the role of being a sandbox in prototyping pre-production networks for DIS-WG applications. As a concrete use case, we should focus on the SRT as it is a prominent alternative to conventional table-based routing for reducing the number of network states and also providing deterministic paths for TE policies.

SRT has recently been made available via Segment Routing by major vendors, but the price tag is still high. This motivates us to use RARE/freeRtr to tackle the issue of cost, but a FOSS solution will be considered valid if only and only if performance metrics behave accordingly under the extremely demanding DIS traffic scenarios. Besides bearing industry-led solutions, recent advances employing a residue number system (RNS) bring a new way of executing fully stateless SRT, in which forwarding decisions rely on a simple modulo operation over a routing label as done in PolKA.

Because much of the state information can be derived from the RNS coding embedded in the packets themselves, this approach could open new pathways to operate and manage networks, making decisions that respond to the progress of data flows, the state of the

end sites, and time-dependent constraints associated with priorities, delivery deadlines, and other policies.

Therefore, on one hand, our research plan will evidently try to provide direct comparisons between segment routing and PolKA regarding performance and functionality metrics. On the other hand, there will also be efforts toward careful experiment design. The latter is important to produce trustworthy realistic scenarios for devising, in the near future and as a followup for our collaboration, a Digital Twin Networks (DTN) structure based on freeRtr for DIS networks. Our initial conjecture is that a DTN architecture for DIS networks can emerge from directly exploiting the following enablers: i) freeRtr's seamless portability from emulation to operation; and ii) RARE/freeRtr's advanced communication interface between the control plane and data plane, which is compliant to P4 Runtime specifications.

Involved Parties

As far as experimental facilities and teamwork are concerned, this proposal also counts on other collaborators, namely, Frédéric Loui at RENATER/GÉANT and Csaba Mate RARE/GÉANT for consulting and granting access to RARE/freeRtr GÉANT's testbed during initial testing for performance validation; and Marcos Schwarz at RNP for advising us on deploying proof-of-principle experiments on the GNA-G's AutoGOLE/SENSE testbed.

- Marcos Schwarz, RNP, marcos.schwarz@rnpbr
- Frederic Loui, Renater, frederic.loui@renater.fr
- Csaba Mate, RARE/FreeRtr, cs@mp.ls
- Harvey Newman, Caltech, newman@hep.caltech.edu
- Frank Slyne, Trinity College Dublin, fslyne@tcd.ie
- Marco Ruffini, Trinity College Dublin, marco.ruffini@scss.tcd.ie
- Eoin Kenny, HEANET, eoin.kenny@heanet.ie
- Qiao Xiang, Xiamen University, xiangq27@gmail.com