SC26 Network Research Exhibition: Demonstration Final Abstract

PATH: Path-aware Adaptive Traffic at the Host for High Performance Data Transfers

Magnos Martinello, Daniel Ventorim, Everson Scherrer Borges, Rafael Guimarães, Cristina K. Dominicini, Moisés Ribeiro, Guilherme Miguel, Jerônimo Bezerra, and Harvey Newman

UFES {magnos.martinello, guilherme.miguel, moises.ribeiro}@ufes.br · IFES {everson, cristina.dominicini, rafaelg, danielv}@ifes.edu.br
FIU {jbezerra@fiu.edu} · Caltech {newman@hep.caltech.edu}

Abstract

Data-Intensive Science (DIS) workflows, such as the Vera C. Rubin Observatory Long-Haul Network, move large datasets continuously over high-latency intercontinental links that cross several Research and Education (R&E) domains. These networks often offer multiple paths with different capacities and delays, yet traffic is typically confined to a single route or spread by Equal-Cost Multi-Path (ECMP) hashing [3], which ignores both path capacity and congestion and allows large flows to collide on the same link.

This demonstration presents **PATH**, a host-controlled approach for **high-performance data transfers** over a stateless network core. In PATH, the Data Transfer Node (DTN) itself selects the end-to-end path of each transfer among a set of pre-computed paths, encoded as RouteIDs, moving it to another path with no reconfiguration of transit nodes. Unlike hash-based host repathing, PATH lets the DTN select the ***complete forwarding path explicitly***: each flow carries a precomputed RouteID, while transit nodes perform only stateless forwarding based on polynomial arithmetic over the binary Galois field GF(2).

PATH builds on PolKA (Polynomial Key-based Architecture) [1], a source-routing scheme in which each core node is identified by a polynomial (nodeID) and each path is encoded as a single route label (RouteID). Every node obtains its output port as the remainder of dividing the RouteID by its own nodeID, so no per-flow or per-path state is kept in the core. At the DTN, programs written in eBPF (extended Berkeley Packet Filter) run in the Linux kernel data path (or in the NIC) and stamp each flow with the RouteID of the chosen path, so that path control stays with the host that owns the transfer [2, 3].

Introduction

At SC26, the demonstration will use PolKA tunnels across the AmLight-ExP and Global P4 Lab (GP4L) infrastructures, terminating at the P4 switch of the Caltech booth, which runs FreeRtr as the routing platform, and at Data Transfer Nodes (DTNs) at the booth and at remote sites.

The contributions demonstrated are twofold:

- **Host-controlled path selection for high-performance transfers:** the DTN assigns each transfer, or each parallel stream of a transfer, to a pre-computed RouteID using eBPF, matching transfers to paths and moving them to another path when needed, with no per-flow state in the core.
- **Explicit-path transport over a stateless core across R&E domains:** PolKA paths spanning AmLight-ExP, GP4L and the Caltech booth show that DTN-controlled routing sustains high-throughput, long-distance transfers over a hybrid core of P4/Tofino and XDP/eBPF nodes.

Goals

1. **Multi-path PolKA deployment:** provision a set of pre-computed PolKA paths with distinct capacities and delays between the SC26 booth and remote sites, and validate stateless GF(2) forwarding on the P4/FreeRtr infrastructure, designed to take part in the Data Tsunami challenge by transporting Pan-American high-performance flows over PolKA.
2. **DTN-driven path selection:** show the source DTN selecting and switching RouteIDs per transfer through eBPF, with no reconfiguration of transit nodes, and measure throughput and transfer completion time on each path.
3. **Host edge with a hybrid stateless core:** demonstrate two data-plane implementations of the PolKA forwarding core, XDP/eBPF software nodes and Tofino hardware nodes, operating on the same paths, with the DTNs acting as the network edge.

Impacts

For UFES (LabNERDS), IFES and the Brazilian National Research and Education Network (RNP), this work addresses a practical need of the RNP e-Science network (https://plataforma.rnp.br/rede-de-e-ciencia), which targets end-to-end transport performance of up to 400 Gbps across heterogeneous R&E domains (GT-ProVip project).

PATH shows that traffic engineering for high-performance data transfers can be controlled from the DTN while the network core remains stateless. Operators can expose the capacity of parallel paths without maintaining per-flow state in transit nodes, and scientific users gain direct control over the path of their bulk transfers, reducing completion times and preventing large flows from colliding on the same link.

SCinet support makes it possible to evaluate host-controlled path selection over real long-distance paths with heterogeneous capacities, high bandwidth-delay products and cross traffic from other exhibitors and production networks. These conditions cannot be faithfully reproduced in a laboratory testbed, and they are precisely where explicit path control at the DTN is expected to pay off.

Future work includes finer-grained distribution at the host (capacity-weighted flow splitting, flowlet-level selection and ECN/CE-aware repathing) within a unified framework to compare these mechanisms [3], offloading the eBPF path-selection logic to SmartNICs and Data Processing Units (e.g., Intel IPU E2000, NVIDIA BlueField), combining path selection with programmable congestion control at the DTN, integrating path computation with orchestration systems such as AutoGOLE/SENSE, and piloting the approach on production paths of RNP and AmLight, in line with ongoing efforts to standardize PolKA.

Resources

Wide-area capacity from trans-continental, international and regional networks (AmLight-ExP, GP4L, RNP and SouthernLight), provisioned through VLAN setup or through AutoGOLE/SENSE. Multiple, preferably disjoint, paths with different capacities between the SC26 Caltech booth and sites in Brazil (Vitória and Ceará) and the United States (Miami) are required to exercise path selection.

P4 devices (Tofino switches) from GP4L and UFES LabNERDS running FreeRtr with PolKA support, including the P4 switches at the Caltech booth, and XDP/eBPF PolKA nodes for the hybrid stateless core.

DTNs from Caltech and UFES LabNERDS running recent Linux kernels with eBPF/XDP (eXpress Data Path) support, plus telemetry collection and a monitoring dashboard for per-path throughput, utilization and transfer completion time.

Involved Parties

- **Caltech HEP:** Harvey Newman (newman@hep.caltech.edu), Raimondas Sirvinskas (raimis.sirvis@gmail.com), Preeti Bhat, Catalin Iordache, Andres Moya, Sravya Uppalapati, Belagam Manaswini.
- **AmLight/FIU:** Julio Ibarra (Julio@fiu.edu), Jeronimo Bezerra (jbezerra@fiu.edu), Vasilka Chergarova (vchergar@fiu.edu), Italo Valcy (idasilva@fiu.edu)
- **GEANT/RENATER:** Frederic Loui (frederic.loui@renater.fr); Csaba Mate, RARE/FreeRtr (cs@mp.ls)
- **RNP:** Marcos Schwarz (marcos.schwarz@rnp.br)
- **SouthernLight/Rednesp:** Ney Lemke (ney.lemke@unesp.br), João Eduardo Ferreira (jef@ime.usp.br), Sandro Rigo (srigo@unicamp.br), Carlos Antonio Ruggiero (toto@rednesp.br), Bruno Baldim (bruno@rednesp.br), Rogério Motitsuki (rogerio@rednesp.br), Geraldo Teodoro da Silva Junior (geraldo@rednesp.br), Paulo Simionato (simionato@rednesp.br) and Rogério L. Iope (rogerio.iope@unesp.br)

References

[1] C. Dominicini *et al.*, "PolKA: Polynomial Key-based Architecture for Source Routing in Network Fabrics," *2020 6th IEEE Conference on Network Softwarization (NetSoft)*, 2020, pp. 326–334, doi: 10.1109/NetSoft48620.2020.9165501.

[2] D. Ventorim *et al.*, "A Path-Aware Routing for Data Intensive Science: Proposal, Deployment and Evaluation in High-Performance Testbed," in *Anais do XXX Workshop de Gerência e Operação de Redes e Serviços (WGRS)*, Natal/RN, Brazil, SBC, 2025, pp. 71–84, doi: 10.5753/wgrs.2025.8769.

[3] D. Ventorim *et al.*,"PATH : Path-aware Adaptive Traffic at the Host for Load Balancing High Performance Data Transfers" to appear on IEEE/ACM International Conference on Utility and Cloud Computing, Florianópolis/SC (UCC 2026).

Acknowledgments

This work was supported by FAPESP/PORVIR-5G (Grant 20/05182-3) and FAPES (Pronem 377/2026). Dr. Martinello is a CNPq fellow (Grant 312058/2023-3). This work was also a recipient of the 2021 *Google Research Scholar Award* in networking and the 2022 Intel *Fast Forward Initiative*.