SC26 Network Research Exhibition: Demonstration Publishable Abstract

StarLight Agentic AIDTN for High-Performance Data/Process Transport with Research Platforms

Chun-Hao Yang, Department of Computer Science, National Yang Ming Chiao Tung University, chyang.cs12@nycu.edu.tw
Jim Chen, International Center for Advanced Internet Research, Northwestern University, jim-chen@northwestern.edu

Abstract

The rapid rise of agentic AI systems—autonomous frameworks that bridge the gap between high-level reasoning and low-level execution—has begun to redefine modern cyberinfrastructure. Building on the DTN-as-a-Service (DTNaaS) lineage , including the foundational work "AIDTN: Towards a Real Time AI Optimized DTN System With NVMeoF,", the StarLight consortium introduces a multi-agent ecosystem that redefines data/process mobility. This architecture deploys distributed autonomous agents to plan, negotiate, and execute end-to-end missions across N × 400+ Gbps WAN infrastructures. Moving beyond static automation, these goal-directed entities employ adaptive reasoning and sophisticated policy to transform high-speed data/process transport into a self-orchestrating, intelligent cyberinfrastructure.

In this paradigm, large scale data/process transport emerges from collaborative agent behavior. Specialized agents continuously interpret telemetry, evaluate constraints, and dynamically reconfigure WAN, LAN, and cloud native resources—including Kubernetes orchestrated environments. Through advanced tool use—such as automated DTN provisioning, protocol selection, congestion avoidance, and real time anomaly mitigation—agents function as distributed operators capable of executing high performance data/process transport strategies at global scale. This shift lays the foundation for autonomous cyberinfrastructure capable of supporting next generation scientific instruments, real time global collaborations, and exascale data/process work flows

The SC26 demonstration will showcase Agentic AIDTN through orchestrating data/process transport services across federated Kubernetes clusters and widely distributed sites. A Jupyter-based control plane allows scientists to collaborate with the Agentic AIDTN collective—auditing agent-generated plans, refining heuristics, and observing real-time self-coordination as the system optimizes Data Transfer Nodes (DTNs) for diverse scientific workflows.

Demonstrated capabilities include Memory-to-Memory and NVMe-over-Fabrics (NVMeoF) transfers using RoCEv2 and TCP across high-radix N × 400+ Gbps WANs. Building on the SC25 achievement of N x 360G+ Gbps aggregate throughput, this year's work positions Agentic AIDTN as the primary engine for an autonomous global data-movement substrate—paving the way for resilient, self-directing infrastructure in the next era of data-intensive science.

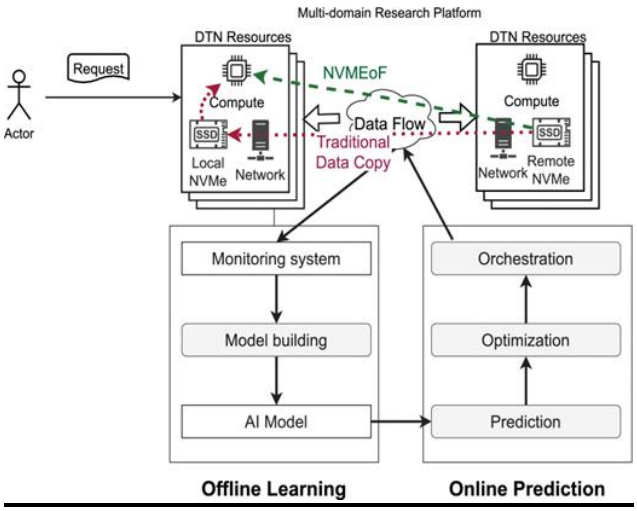


Figure 1. AIDTN Architecture

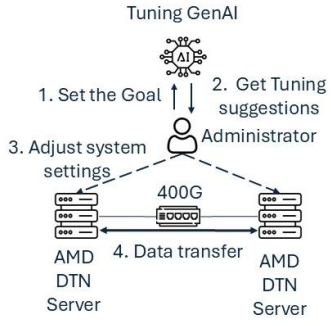


Figure 2. GenAI System Architecture

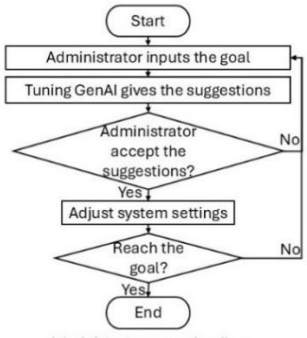


Figure 3. Tuning GenAI Tuning Administrator operation flow

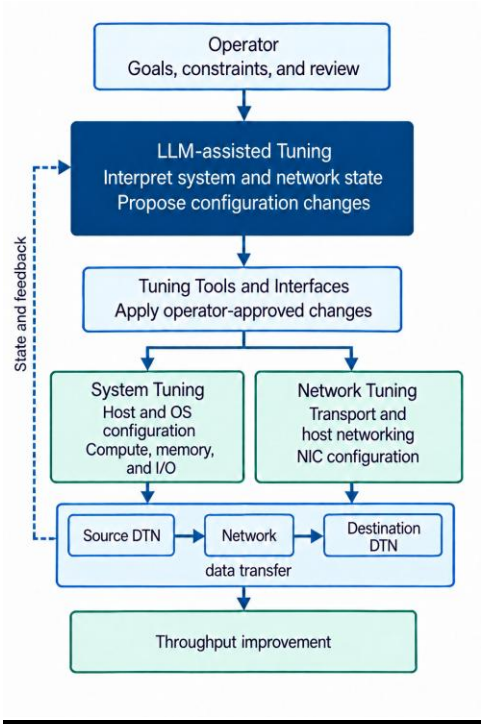


Figure 4. LLM-based System and Network Tuning Architecture

Goals

The overarching goal is to transition global data infrastructure from manual/static automation to a self-orchestrating, autonomous substrate. Here is a breakdown of the specific objectives this project accomplishes:

1. Autonomous Lifecycle Management of Data/Process Transport: The primary goal is the creation of Agentic AIDTN, a system where human operators no longer need to manually tune parameters. Instead, distributed agents handle the entire lifecycle—planning, negotiating, and executing—of end-to-end data/process movement across massive pipes (N x 400+ Gbps).

2. Intelligent, Dynamic Resource Reconfiguration

The project moves beyond "set it and forget it" scripts. The goal is to achieve real-time infrastructure adaptation where agents:

- Continuously interpret telemetry.
- Dynamically reconfigure WAN, LAN, and Kubernetes resources.
- Automatically select protocols and mitigate anomalies (like congestion) without human intervention.

3. Scaling "Data/Process" Mobility to Exascale

By integrating NVMe-over-Fabrics (NVMeoF) with TCP and RoCEv2, the project aims to prove that agentic AI can handle the most demanding hardware stacks. The goal is to move past the SC25 benchmarks and demonstrate that these agents can manage $N x 400+ Gbps throughput at a national/global scale.

4. Establishing a Human-Agent Collaborative Framework

A key goal is the "Human-in-the-loop" model via a Jupyter control plane. This accomplishes a balanced partnership where scientists aren't replaced, but rather "collaborate with the collective" auditing plans and refining workflows while the AI handles the low-level, high-speed execution.

Impacts

Foundation for Future Scientific Research Infrastructure: Ultimately, the project intends to provide a prototype for next-generation cyberinfrastructure. It seeks to prove that multi-agent systems are one the viable ways to support the extreme data demands of:

- Next-gen scientific instruments (e.g., giant telescopes or particle accelerators).
- Real-time global collaborations.
- Exascale workflows that are too complex for traditional research workflows

Resources

SC26 SCinet WAN services totaling 1.2+ Tbps end-to-end are required. This includes connectivity from the StarLight International/National Communications Exchange Facility in Chicago to the StarLight booth at the SC26 venue, between StarLight and the Joint Big Data Testbed (JBDT) Facility in McLean, Virginia, between the JBDT Facility and the StarLight booth at the SC26 venue, and among all sites. Additionally, the 400 Gbps ESnet testbed at Berkeley, connected to the StarLight Facility, will be utilized.

Involved Parties

- Jim Chen, iCAIR, jim-chen@northwestern.edu
- Se-Young Yu, ESnet, youf3@es.net
- Ting-Yu Lin, NCHC/NIAR, 2503115@niar.org.tw
- Chun-Hao Yang, NYCU, chyang.cs12@nycu.edu.tw
- Fei Yeh, iCAIR, fyeh@northwestern.edu
- Joe Mambretti, iCAIR, j-mambretti@northwestern.edu
- Marc Lyonnais, Ciena, mlyonnai@ciena.com
- Gaurauvdeep Shami, Ciena, gshami@ciena.com
- Meng-Hsun Tsai, NYCU & NCHC/NIAR, tsaimh@cs.nycu.edu.tw