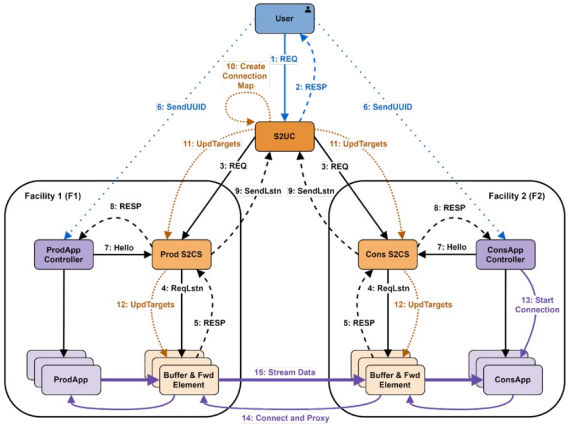
4th GLOBAL RESEARCH PLATFORM WORKSHOP

Argonne NATIONAL LABORATORY

SciStream: Architecture and Toolkit for Data Streaming between Federated Science Instruments



Rajkumar Kettimuthu
Argonne National Laboratory and The University of Chicago

With contributions from Joaquin Chung et al.

Limassol, Cyprus, Oct. 10 2023

Data Processing in Light Source Facilities

A Science Driver

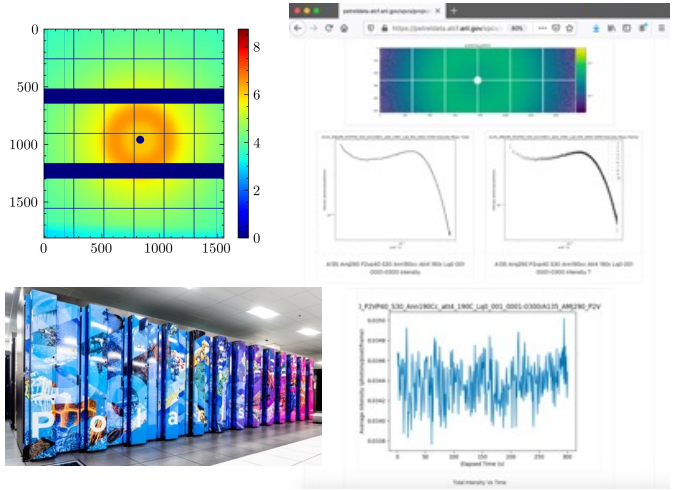
Experiment

Transfer data

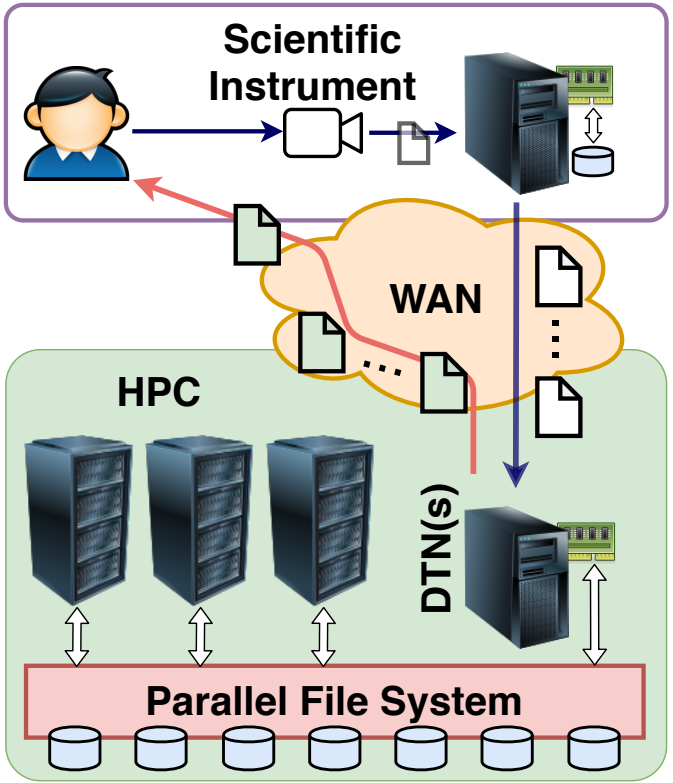
Data Analysis

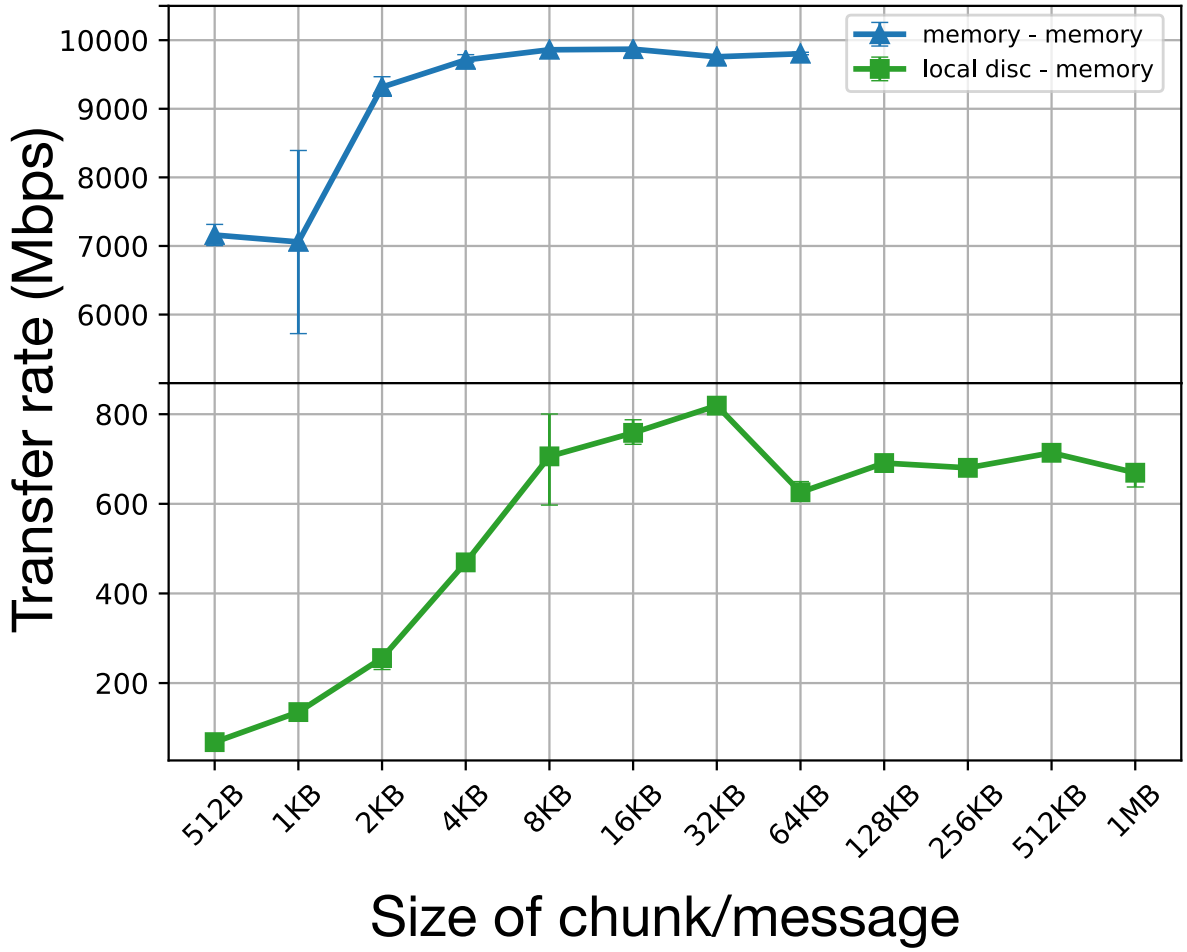
Decision/
Steer

Traditional File-based Data Movement

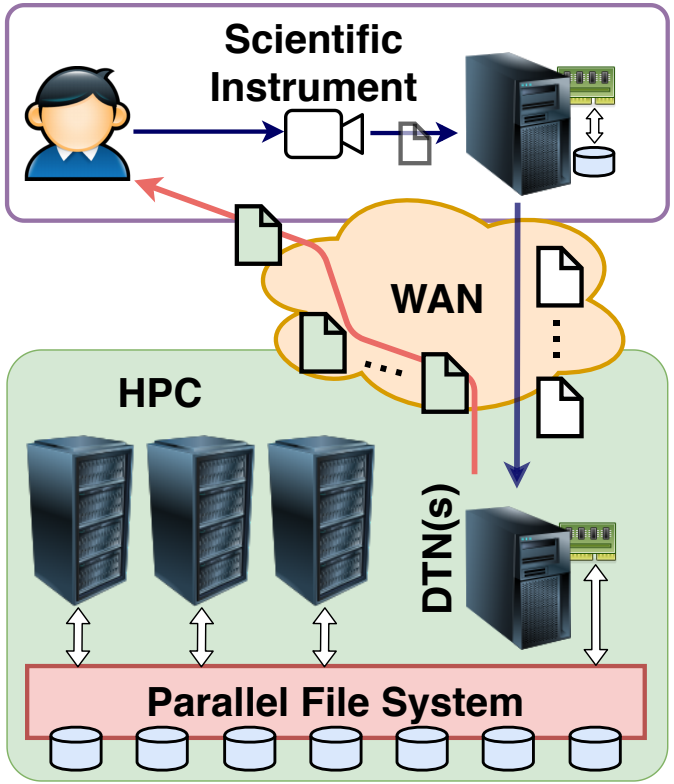


Speckle data from the APS 8-ID-I beamline (top left) is automaticity transferred to the Polaris supercomputer (bottom left) where it is processed on-demand and displayed in a Globus web portal (right).

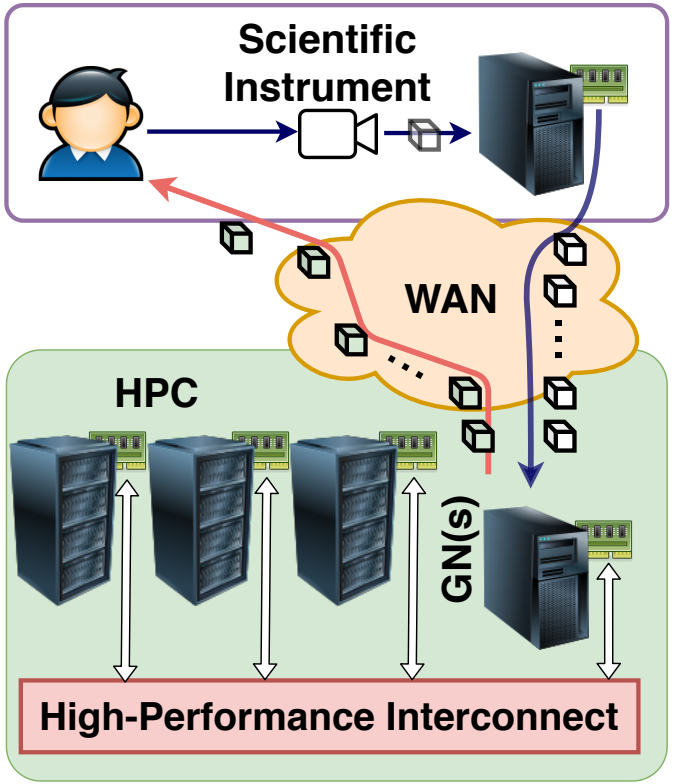




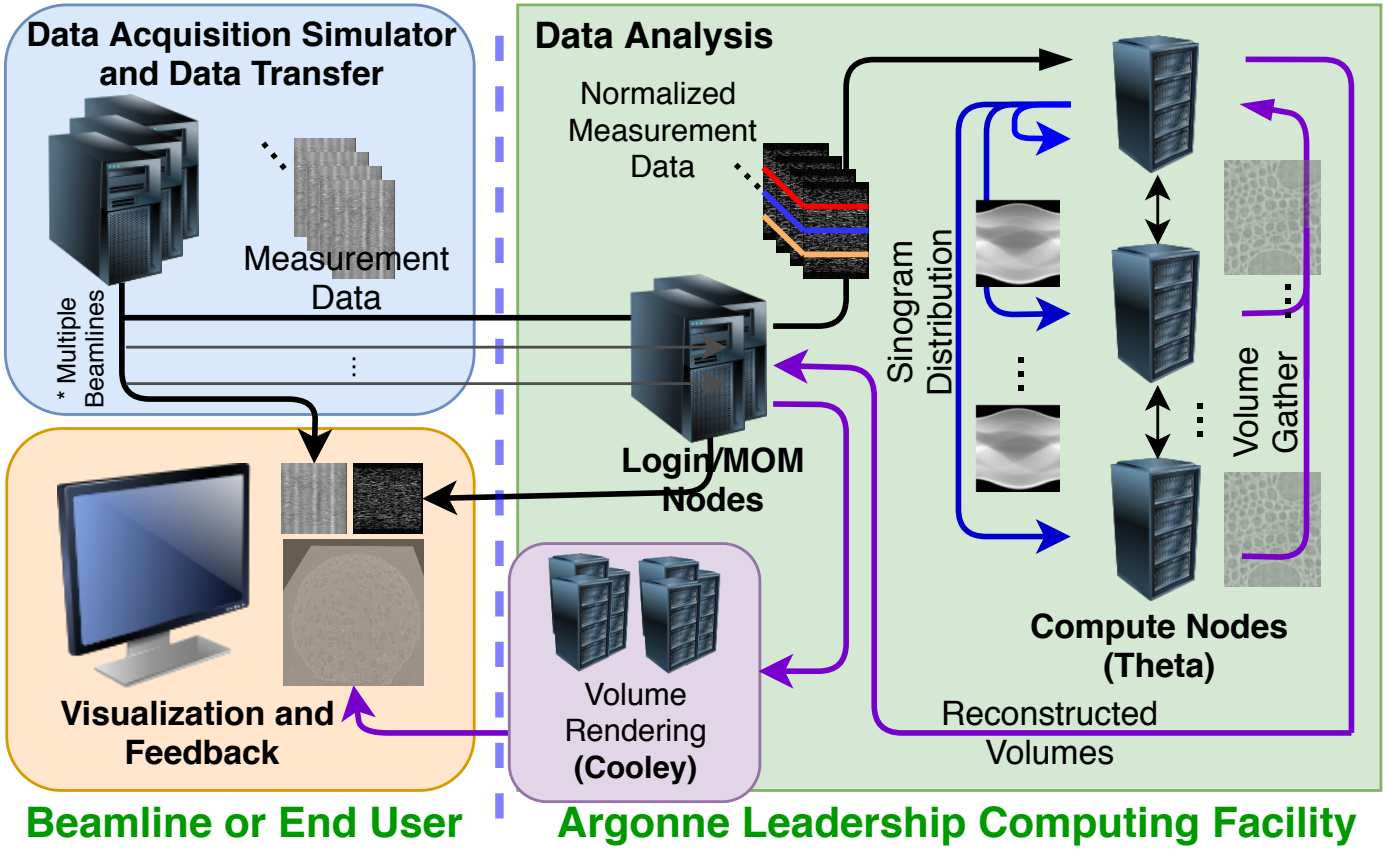
Traditional File-based Data Movement



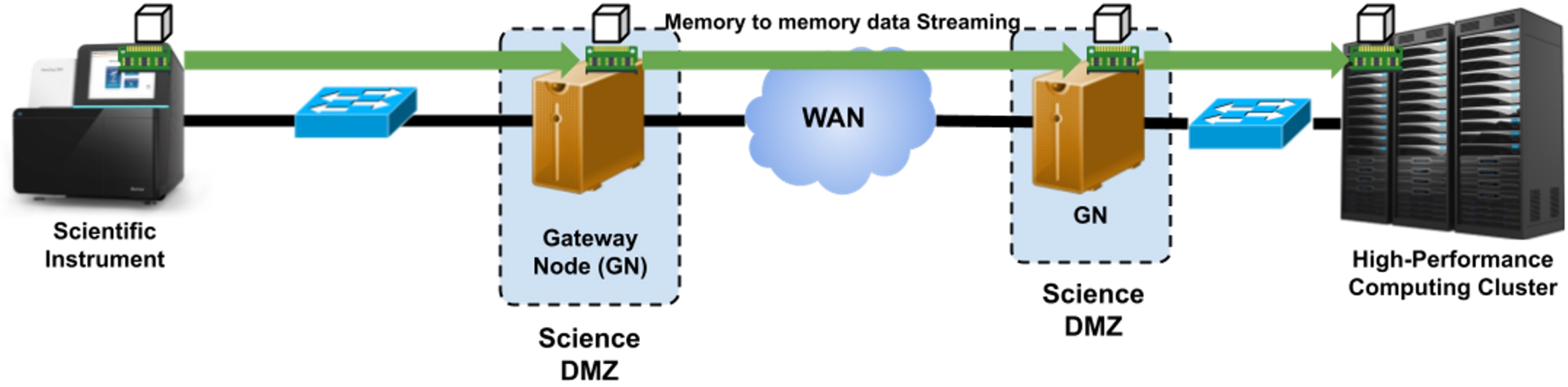
Memory-to-Memory streaming from instrument to HPC



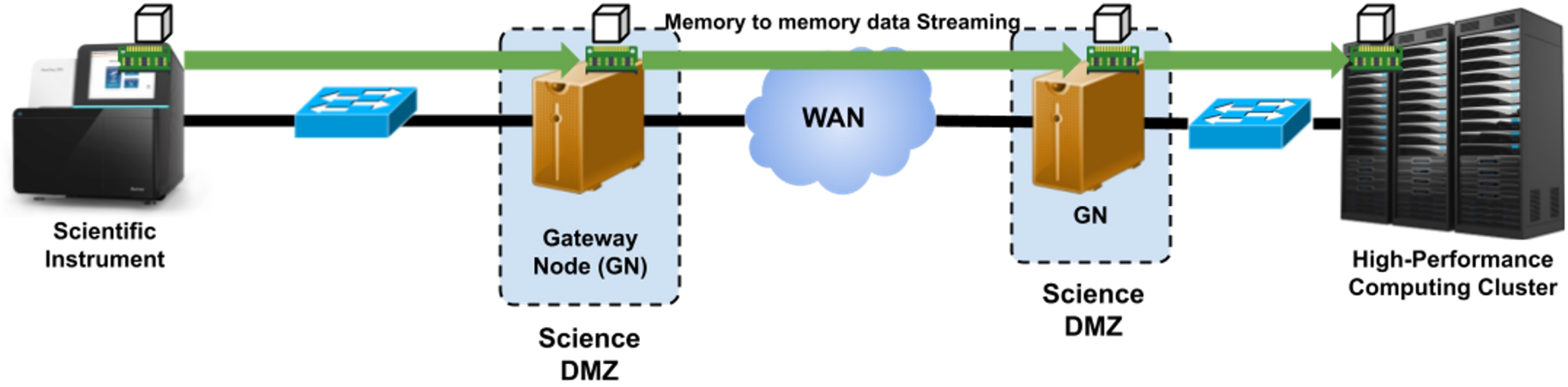
SC19 Tech Challenge: Real-time Stream Analysis over WAN



Multiple Connections in the End-to-End Path



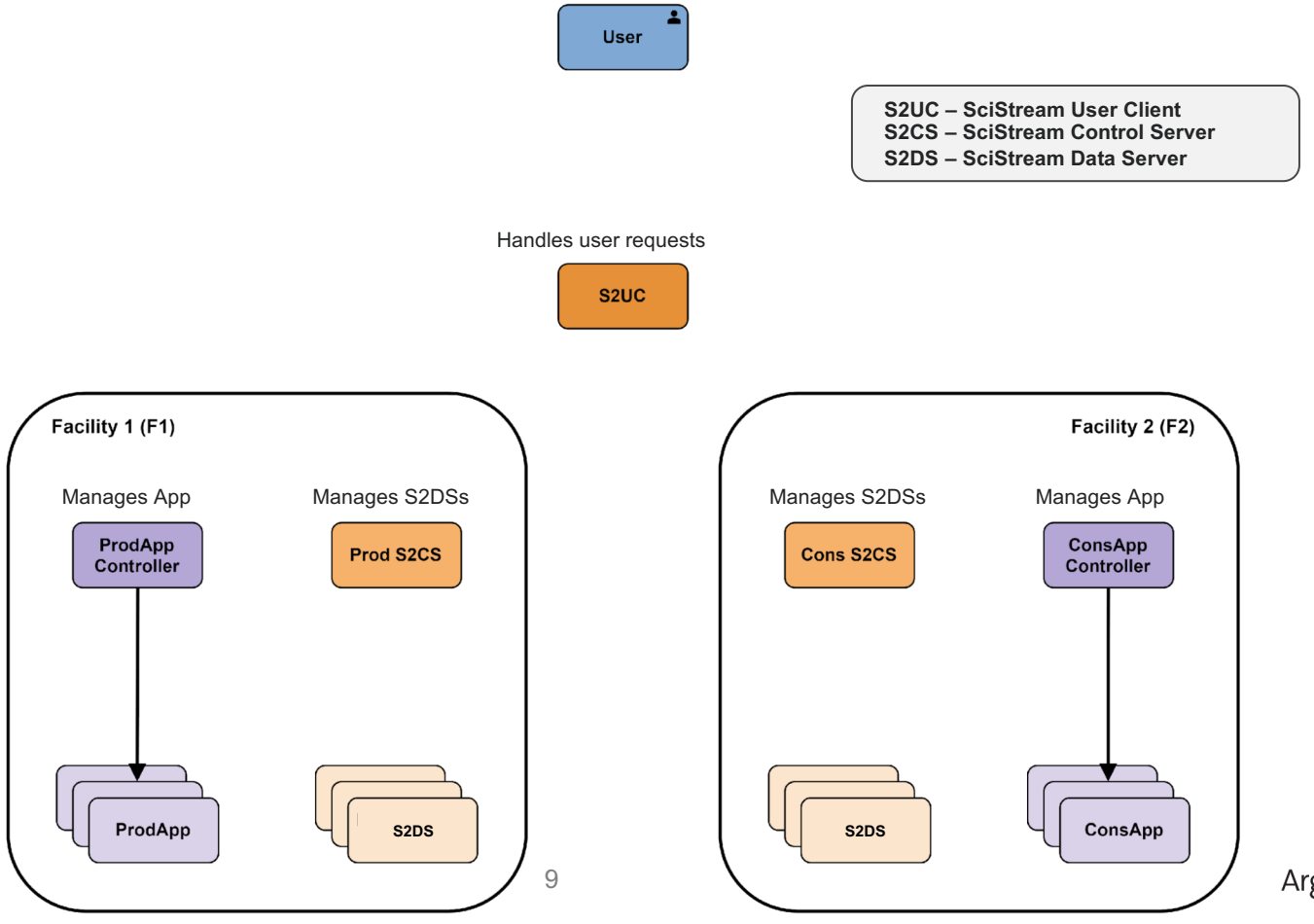
Multiple Connections in the End-to-End Path



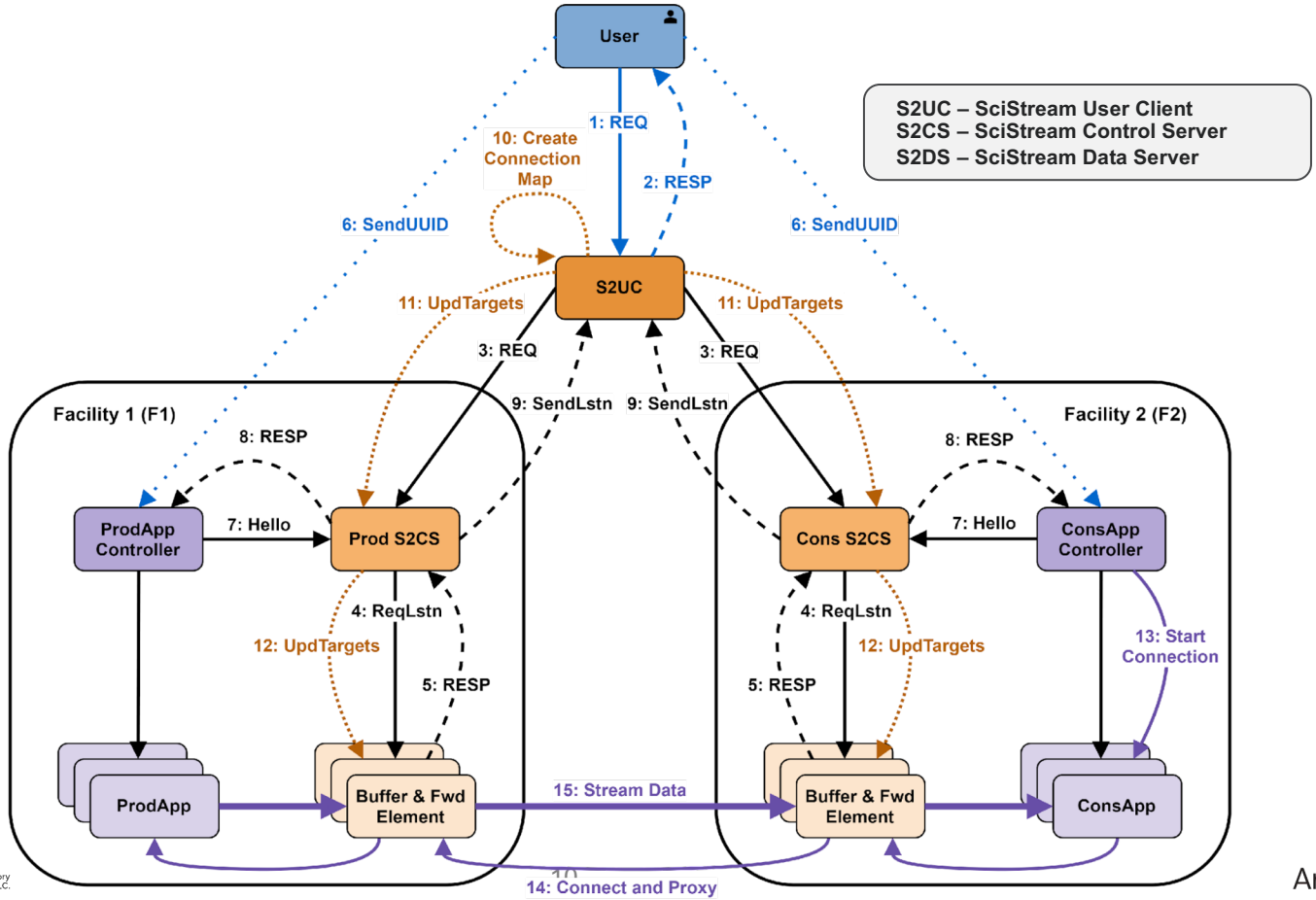
Design considerations

- Third-party Streaming
- Secure Streaming
- General and Transparent Streaming
- Provisioned vs. Best-effort resources

SciStream Components



SciStream Protocol



SciStream Implementation

SciStream Control Server (S2CS) and Data Server (S2DS)

S2CS:

- Implemented in Python using state machine
- Memory footprint is 10MB, data streaming request is completed in ~0.12 s while a release is completed in 0.003 s

S2DS (Implementation options):

- L3 NAT or tunnels
- L4 Proxy (TCP or UDP)
- L7 (Application) Proxy

SciStream Implementation

SciStream Control Server (S2CS) and Data Server (S2DS)

S2CS:

- Implemented in Python using state machine
- Memory footprint is 10MB, data streaming request is completed in ~0.12 s while a release is completed in 0.003 s

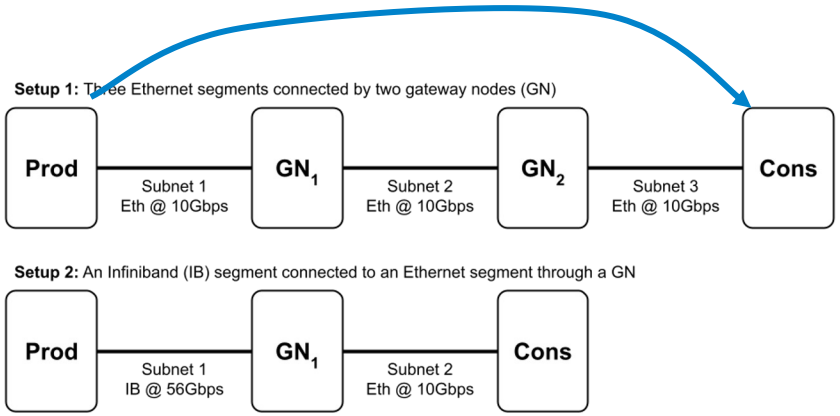
S2DS (Implementation options):

- L4 Proxy (TCP or UDP)
  - Buffer location – user space vs kernel vs NIC
  - DPUs

SciStream Evaluation

S2DS Implementation Options

Experimental setup on Chameleon



Methodology: compare streaming over SciStream against the ideal case where producer and consumer have direct connectivity.

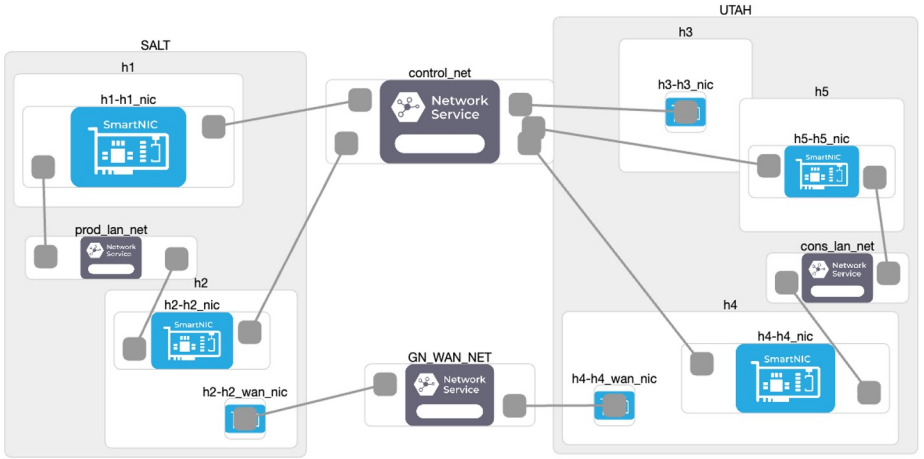
Measurements:

- Goodput evaluation
- Added latency and intermessage delay variation

Experimental Setup on FABRIC

Topology diagram drawn by FABRIC's GUI

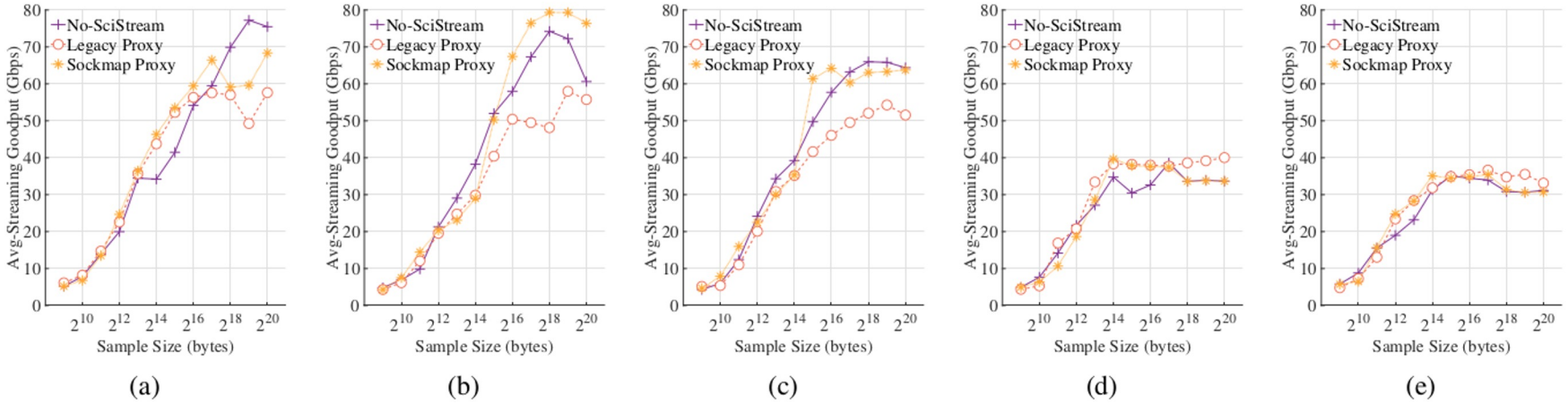
- We request FABRIC resources from two sites and connect them via a WAN link and a separate control network
  - On each site we request two compute nodes connected by a LAN
  - All compute nodes have 100GbE NICs
  - One site has an extra compute node for running SciStream control protocol
- We conduct experiments in five setups with different values of RTT in the WAN



	LAN	Metro	Short WAN	WAN	Long WAN
Prod LAN	0.087	0.092	0.167	0.165	0.161
Cons LAN	0.100	0.105	0.179	0.148	0.160
GN WAN	0.253	5.293	23.998	57.848	143.370
Overall	**0.440**	**5.490**	**24.344**	**58.161**	**143.691**

Evaluation Results

Average TCP Streaming Goodput



Goodput performance of two TCP implementations of SciStream S2DS (legacy and sockmap proxies) compared to the No-SciStream case over five network setups: (a) LAN, (b) Metro, (c) Short WAN, (d) WAN, and (e) Long WAN.

Demo

https://drive.google.com/file/d/1oT9KZKmWsQfPwASAsQcTLD2Jw-5rhsvg/view?usp=sharing

Acknowledgments

U.S. DEPARTMENT OF ENERGY Argonne National Laboratory is a U.S. Department of Energy laboratory managed by UChicago Argonne, LLC.

Argonne NATIONAL LABORATORY

Questions

Thanks!

Questions: kettimuthu@anl.gov

Acknowledgments