SC23

Denver, CO | i am hpc.

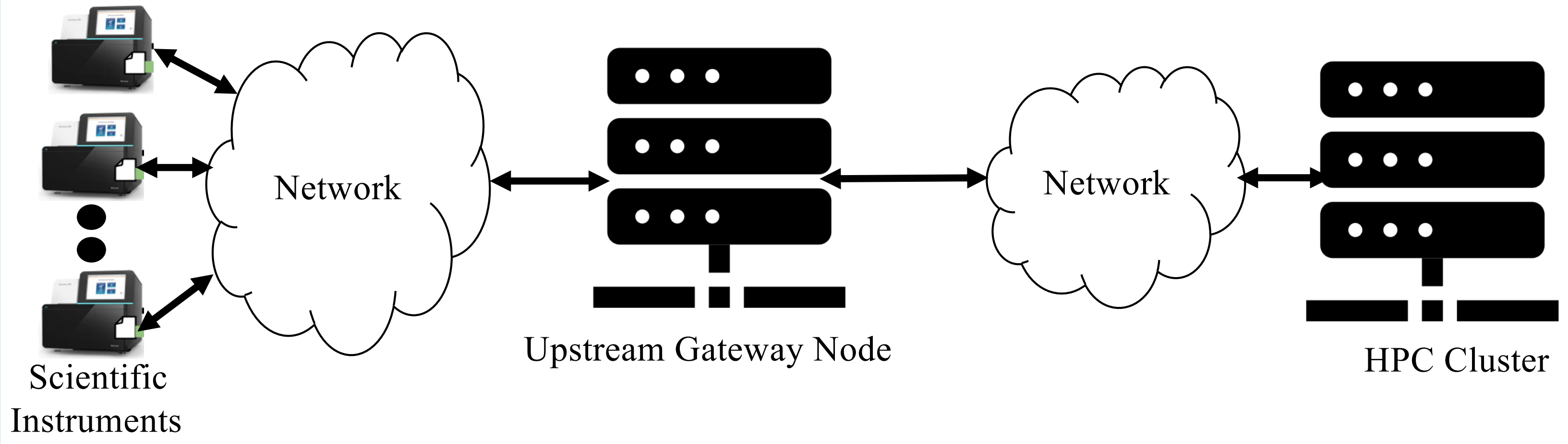
Throughput Optimization With a NUMA-aware Runtime System for Efficient Scientific Data Streaming

Hasibul Jamil, Joaquin Chung, Tekin Bicer, Tevfik Kosar, Rajkumar Kettimuthu

University at Buffalo The State University of New York

Argonne NATIONAL LABORATORY

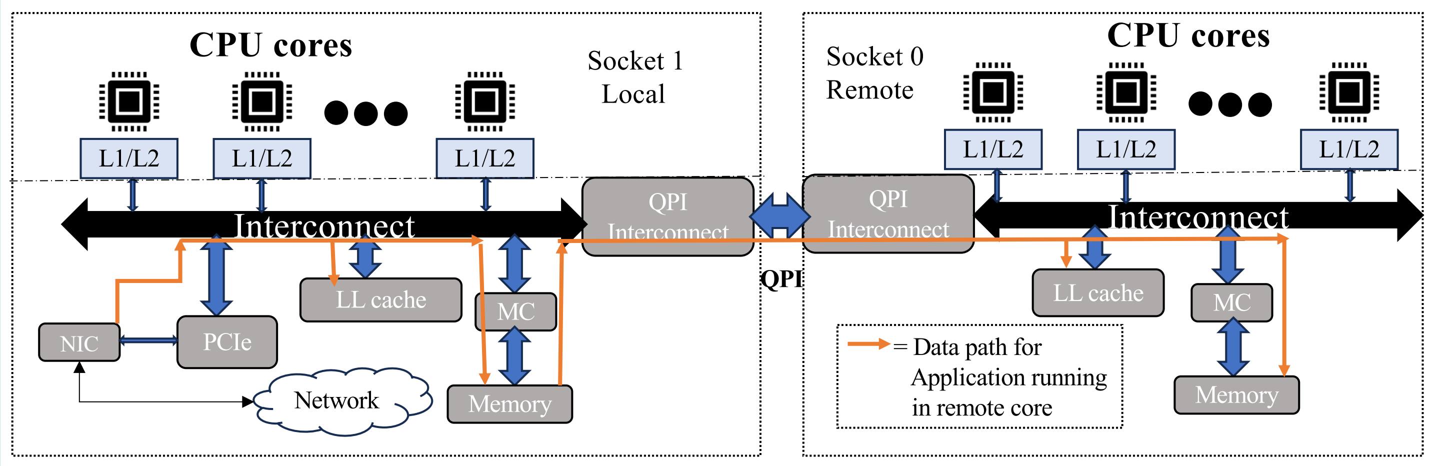
Addressing High-Speed Data Streaming in Scientific Research



- Rapid Data Generation Rates
- Infrastructure Bottlenecks
- Need for Upgraded Upstream Processing:

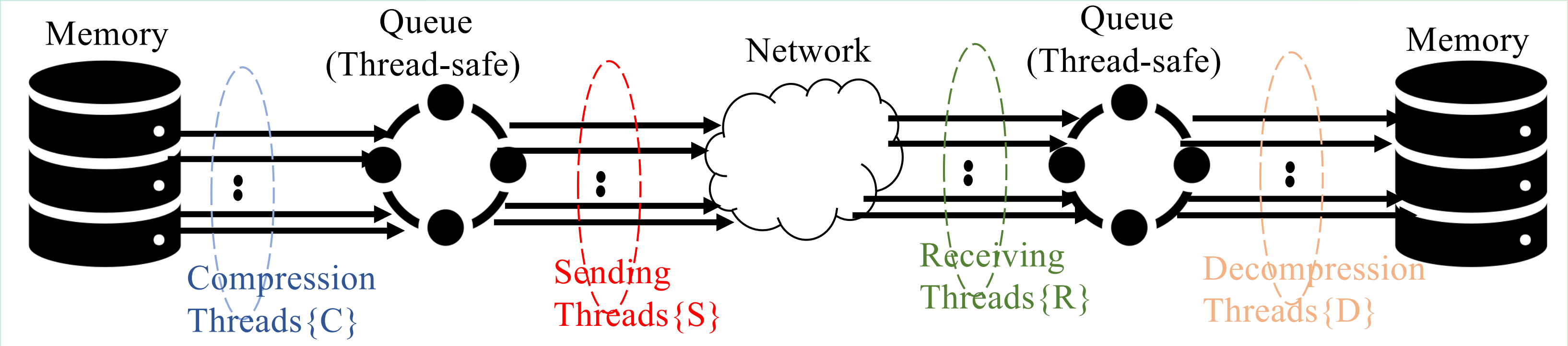
- Gateway Node Functions
- Optimized System Architecture
- Scalability for Future Demands

NUMA Considerations and Performance Management



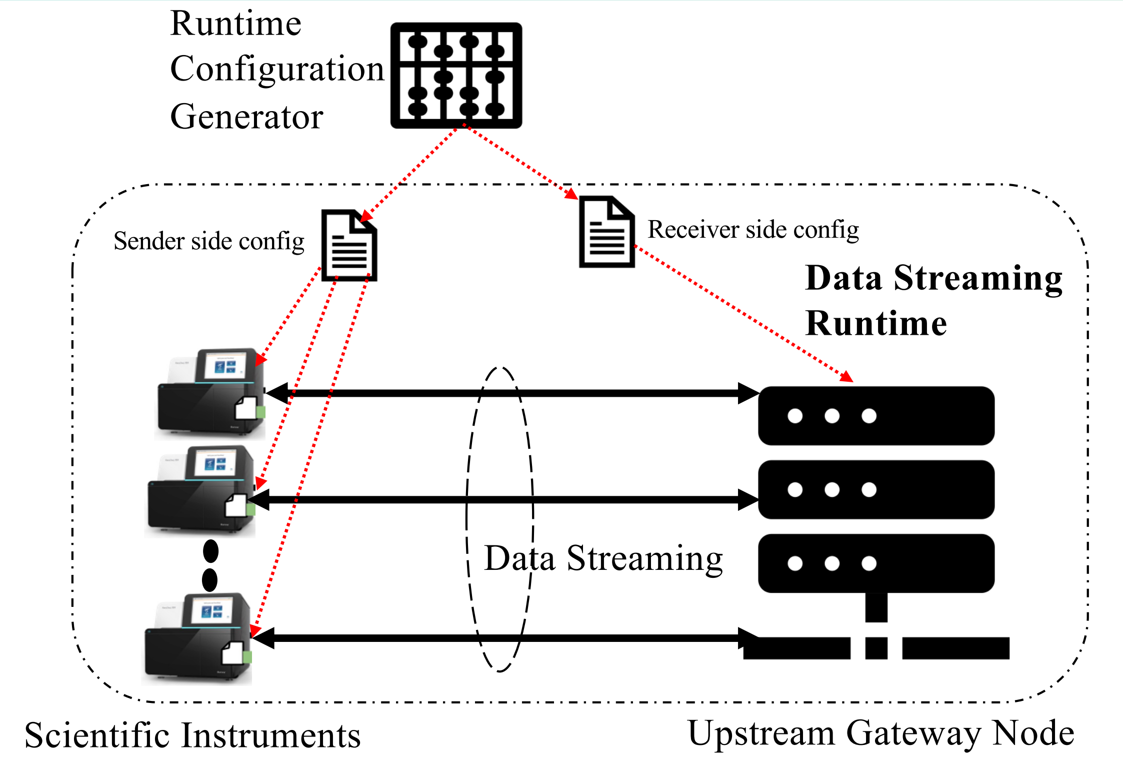
- NUMA Architecture Basics
- Memory Access in NUMA
- NIC Operation and NUMA

The Role and Objectives of the Runtime System



- Optimized Packet Processing
- Reducing Cross-Socket Traffic

Overview of the Runtime System Framework



- Runtime Configuration Generator
- Distributed Framework

Dataset and Compression-Decompression Algorithms

- **Dataset Characteristics:**
  - Utilized a synthesized 16 GB dataset reflective of real tomographic data, processed in 11.0592 MB chunks.

- **Compression-Decompression Algorithms:**
  - LZ4 algorithm selected for its speed and favorable compression ratio, achieving an average 2:1 compression.

Compression Behavior and Performance with NUMA

Goal: Maximize Resource Utilization and Minimize Network I/O

- Use available CPU cores for efficient data compression, effectively doubling the data transfer speed.

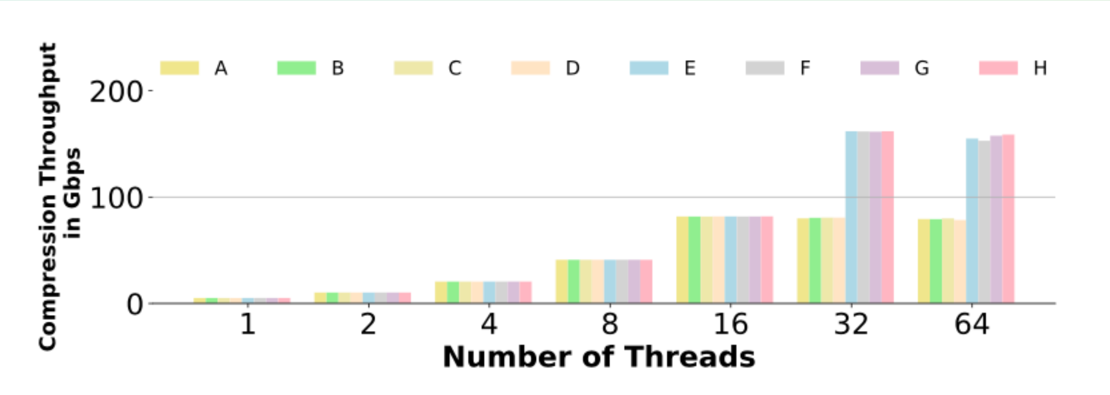
Strategy: Employ Data Compression to Enhance Throughput

- Implement LZ4 compression algorithm for real-time data compression with a 2:1 compression ratio.

Observation: Compression Throughput and CPU Core Count

- Increased thread count improves compression speed up to the number of available CPU cores; beyond that, performance plateaus due to context switching.

Configuration	Memory Domain	Execution Domain
A	0	0
B	0	1
C	1	0
D	1	1
E	0	0 & 1
F	1	0 & 1
G	0	OS
H	1	OS



Decompression Behavior and Performance with NUMA

Goal: Analyze Decompression Speed Influencers

- Determine the impact of the number of decompression threads and their NUMA domain alignment on performance.

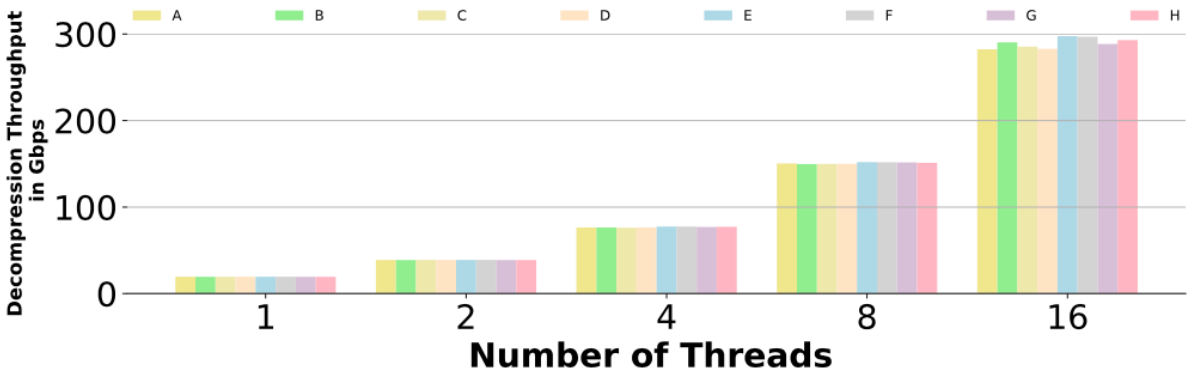
Strategy: Optimize Thread Distribution Across NUMA Domains

- Decompression speed improves with additional threads, with best performance when evenly spread across NUMA domains.

Observation: Decompression Throughput Unaffected by NUMA Domain

- Decompression performance remains consistent regardless of the NUMA domain of data storage or execution.

Configuration	Memory Domain	Execution Domain
A	0	0
B	0	1
C	1	0
D	1	1
E	0	0 & 1
F	1	0 & 1
G	0	OS
H	1	OS



University at Buffalo The State University of New York

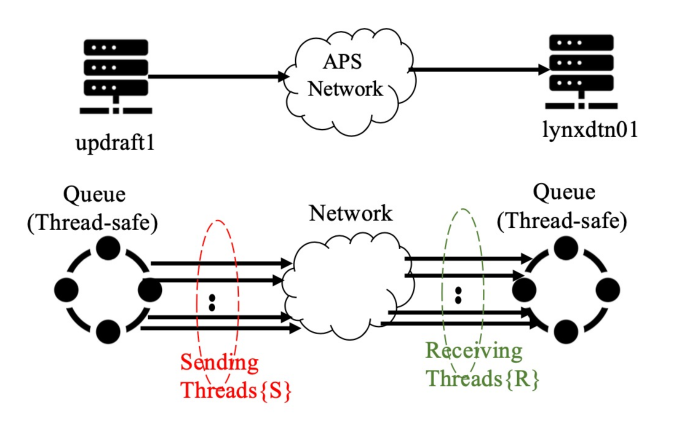
Argonne NATIONAL LABORATORY

Sending and Receiving Threads Performance with NUMA

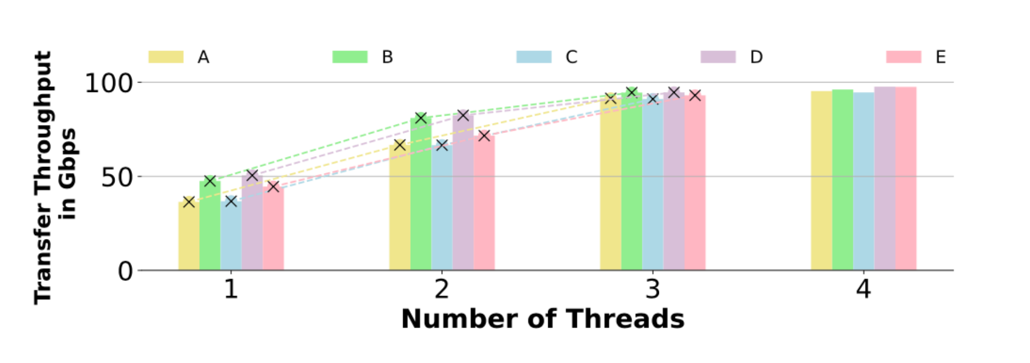
- **Goal: Understand Thread Influence on Network Throughput**
  - Examine the effect of the number and location of sending and receiving threads on network throughput.

- **Strategy: Symmetrical Thread Arrangement Across NUMA Domains**
  - Deploy an equal number of sending and receiving threads, creating a balanced TCP streaming environment.

- **Observation: Receiving Thread Location Impacts Throughput**
  - Placing receiving threads in the same NUMA domain as the NIC significantly boosts throughput, especially for smaller thread counts.



Configuration	Sender Socket	Receiver Socket
A	0	0
B	0	1
C	1	0
D	1	1
E	OS	OS

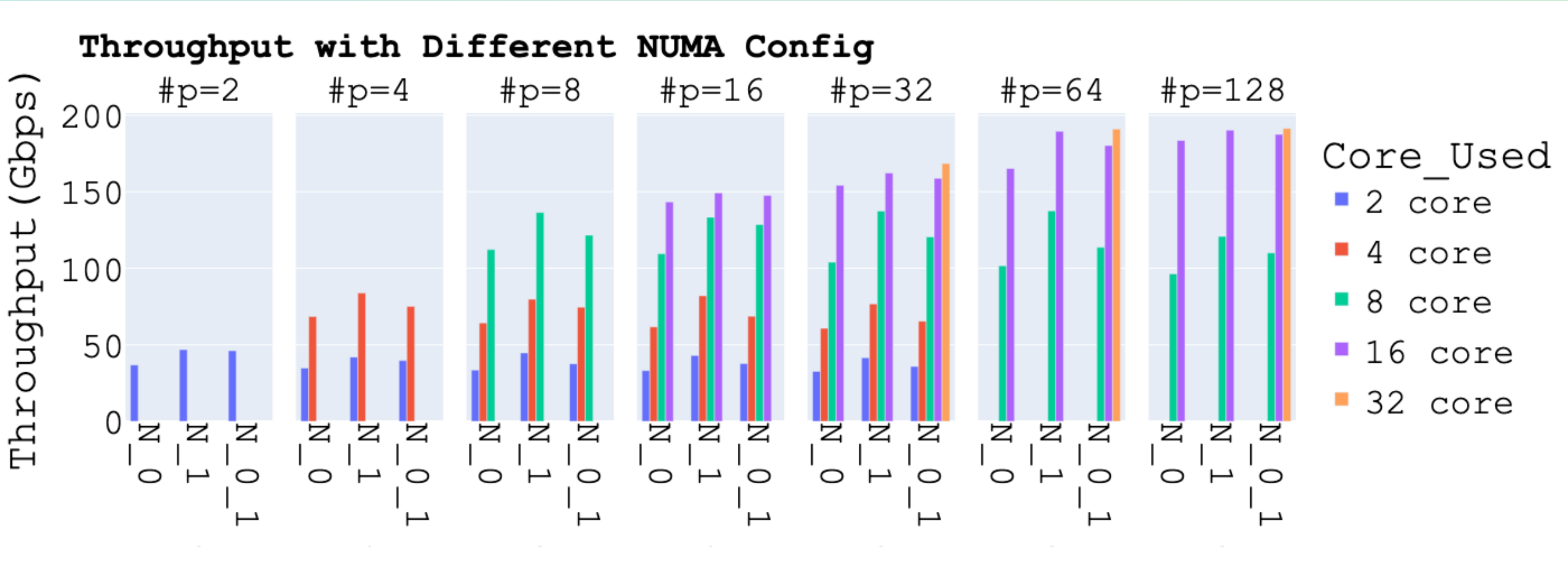


Network Performance and NUMA

Goal of the Experiment : Investigate network transfer throughput and core affinity on data streaming between facilities with high-bandwidth connections.

Strategic Use of NUMA: Utilize NUMA-aware strategies to improve throughput by assigning tasks to cores that have local memory access to the NIC.

Observations from the Experiment:



University at Buffalo The State University of New York

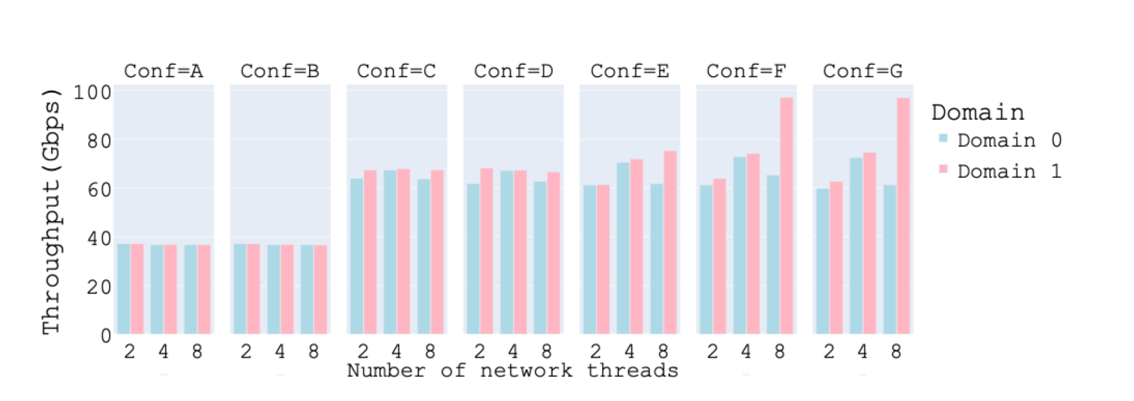
Argonne NATIONAL LABORATORY

Single Stream Evaluation in Runtime System

- **Goal: Assess Runtime System Efficiency with a Single Data Stream**
  - Evaluate system performance across various configurations for compression, decompression, and transmission-reception threads.

- **Strategy: Diverse Thread Configuration Experiments**
  - Use two interconnected machines capable of 100 Gbps transfers to test different combinations of thread counts and execution domains.

Configuration	#of compression Threads	#of decompression Threads
A	8	4
B	8	8
C	16	8
D	16	16
E	32	4
F	32	8
G	32	16



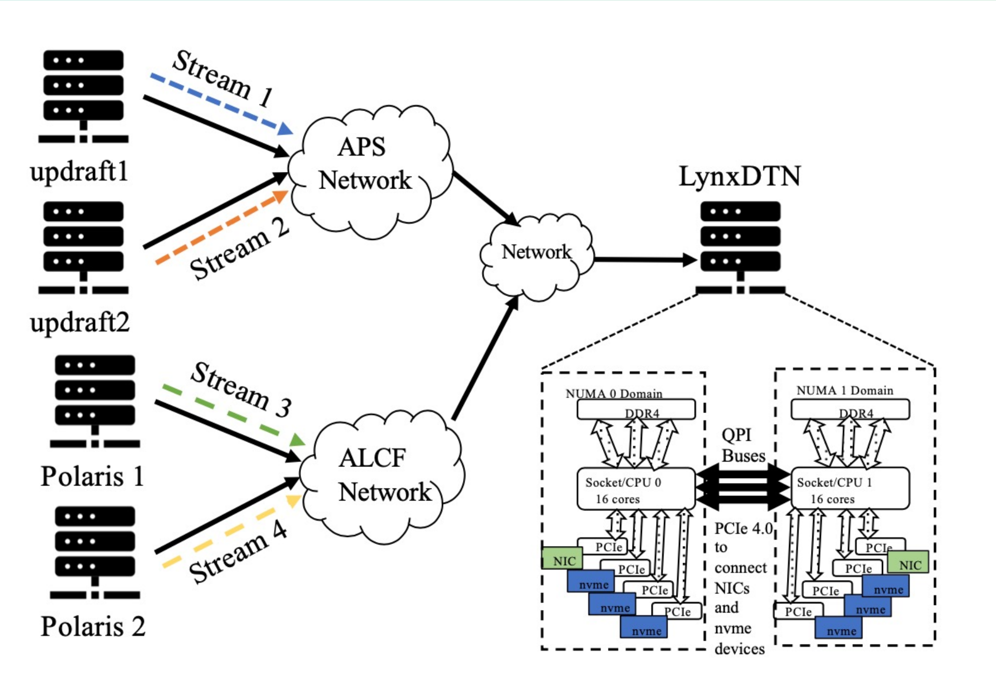
- **Observation: Bottlenecks and Throughput Efficiency**
  - Throughput varies with the number of compression threads; end-to-end performance peaks with receiver threads in NUMA domain 1, achieving 97 Gbps in optimal settings.

University at Buffalo The State University of New York

Argonne NATIONAL LABORATORY

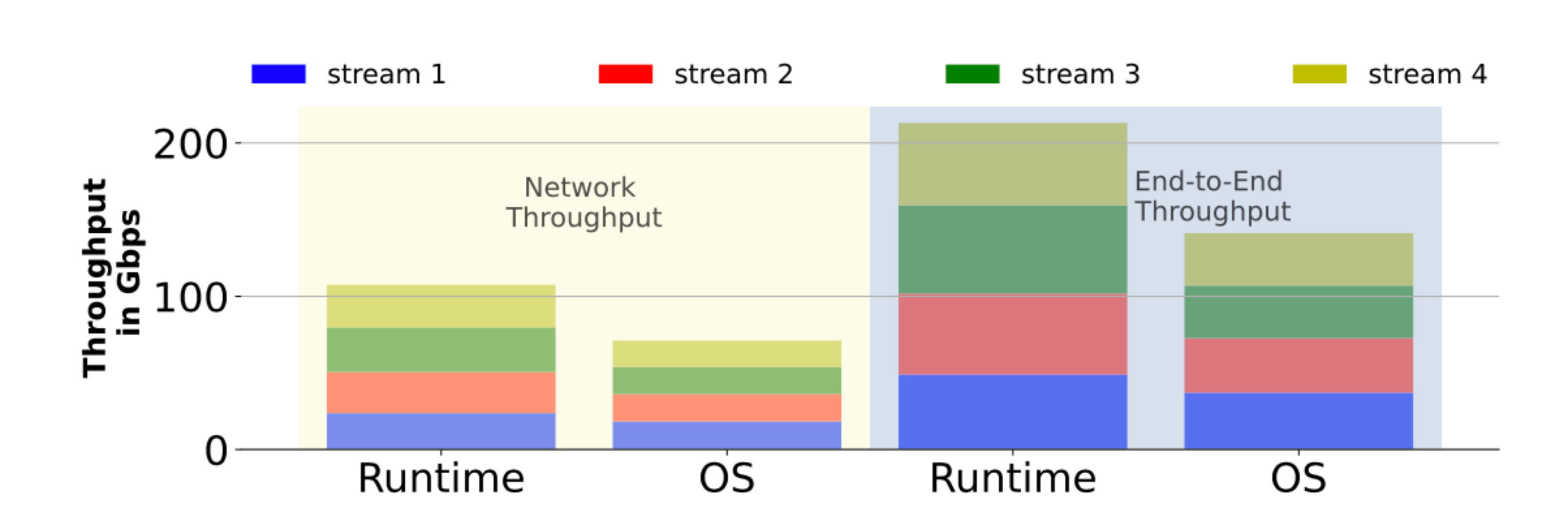
Multi Stream Evaluation in Runtime System

- **Goal: Compare Runtime System and OS-Determined Thread Placement**
  - Test the runtime system's effectiveness against an OS-controlled thread execution location strategy.
- **Strategy: Multi-Source Data Stream Generation and Reception**
  - Generate four concurrent data streams across machines with varying architectures, assessing combined and individual network and end-to-end throughput.



Multi Stream Evaluation in Runtime System

- **Observation: Runtime System Superiority in Throughput**
  - The runtime system, leveraging detailed architectural knowledge, significantly outperforms the OS's autonomous thread placement, achieving 105.41 Gbps network and 212.95 Gbps end-to-end performance.



Conclusion - Optimizing Data Streaming with NUMA-Aware Runtime System

- Comprehensive System Evaluation
- NUMA Optimization Proven Effective
- Multi-Stream Performance Superiority
- Single Stream Insights
- Empirical Evidence of Efficiency
- Future-Proofing Data Transmission

Future directions

- **Towards Dynamic Pinning:**
  - Current system utilizes static CPU pinning which, while effective, does not adapt to fluctuating workloads in multi-user environments.

The project's GitHub repository : https://github.com/H-jamil/ha4hpdt.git.

Questions:

mdhasibu@buffalo.edu

Acknowledgments: DE-AC02- 06CH11357

U.S. DEPARTMENT OF ENERGY | Office of Science

University at Buffalo The State University of New York

Argonne NATIONAL LABORATORY

Extra