Network-Induced Memory Contention

Taylor Groves, Ryan Grant, and Dorian Arnold

UNM

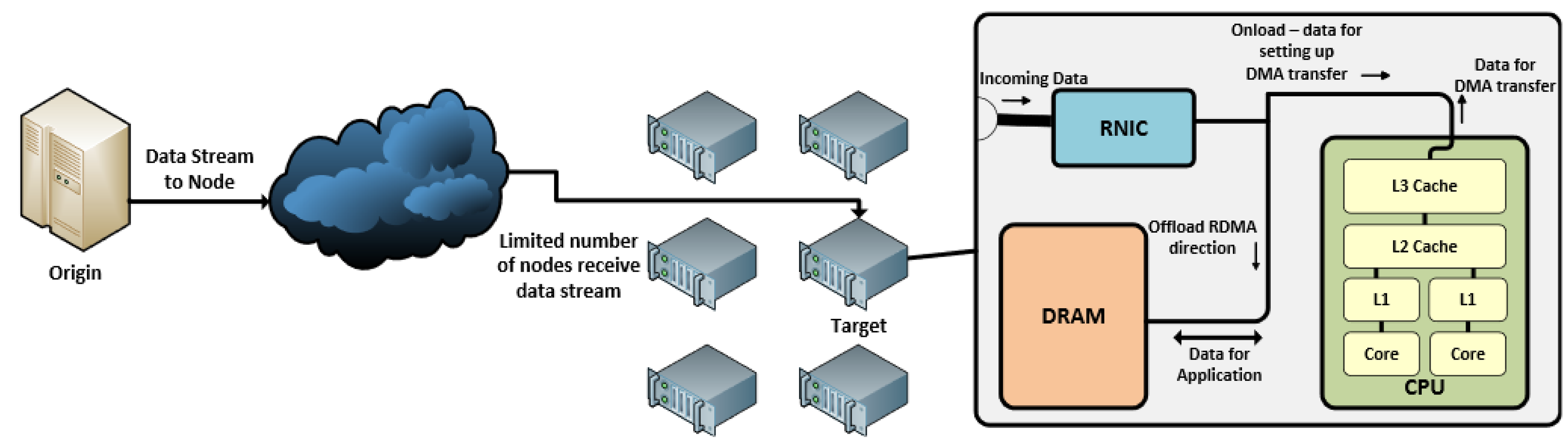
Sandia National Laboratories

What is Network-Induced Memory Contention (NiMC)?

Out-of-band network operations, e.g. RDMA, that adds pressure to the memory subsystem

Exascale solutions increasingly rely on RDMA

- Asynchronous Many Task Applications
- Partitioned Global Address Space Applications
- Data Staging
- In-Situ Analytics



NiMC Performance Impact Case Studies

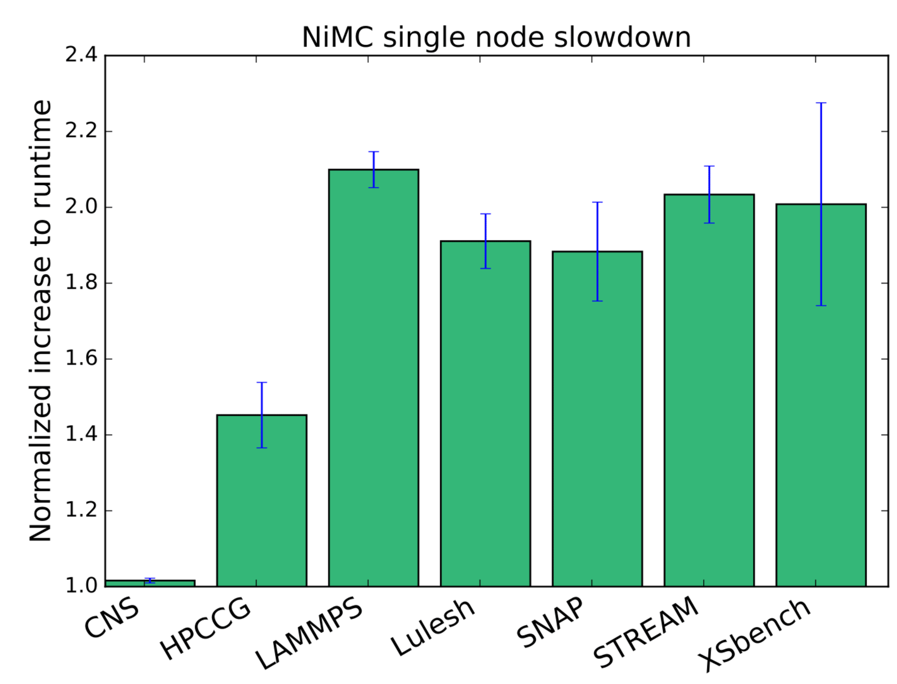
Single-node Study

- RDMA writes at 2-5 GBps
- Memory & cache contention observed on six of eight test platforms
- Worst impact on system with onload NICs

LAMMPS Multi-node Study

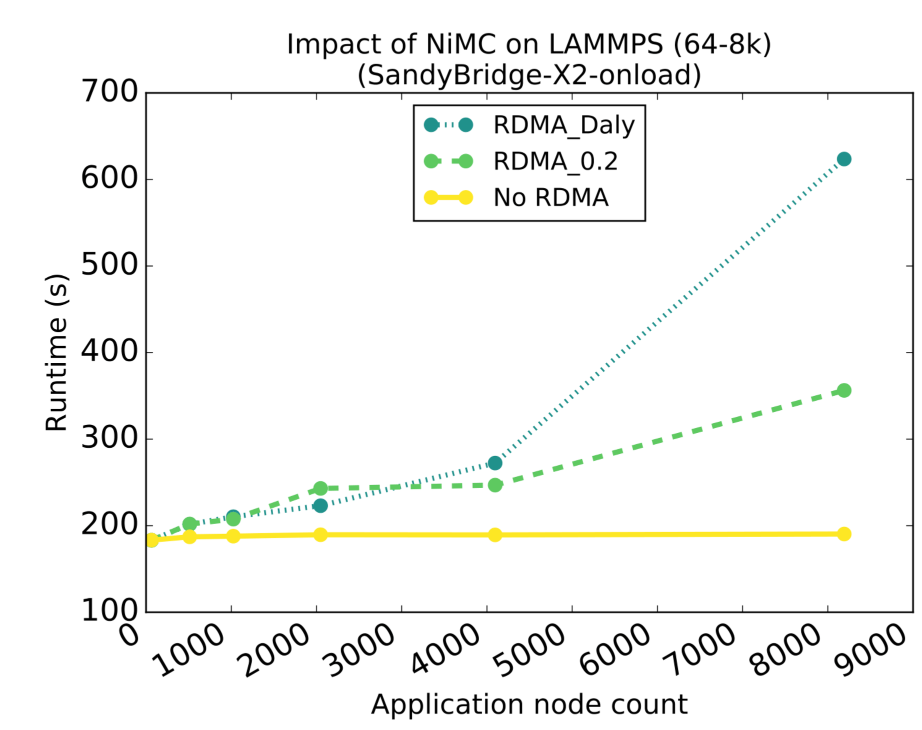
- 2X Sandybridge system with onload NICs
- 1 Sec. RDMA writes on a subset (0.2-0.5%) of nodes
- > 3X slowdown (LAMMPS)

NiMC on a Single Node



NiMC can significantly slow down single node runs by more than 100%

NiMC across Multiple Nodes

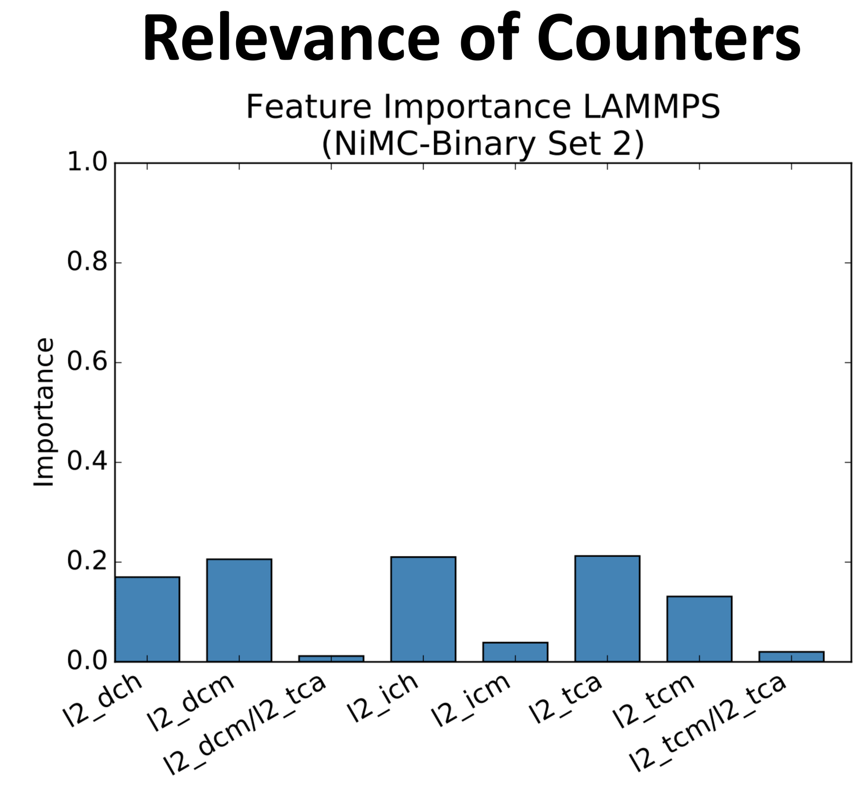


NiMC increases LAMMPS runtime by 3X

Detection and Diagnosis

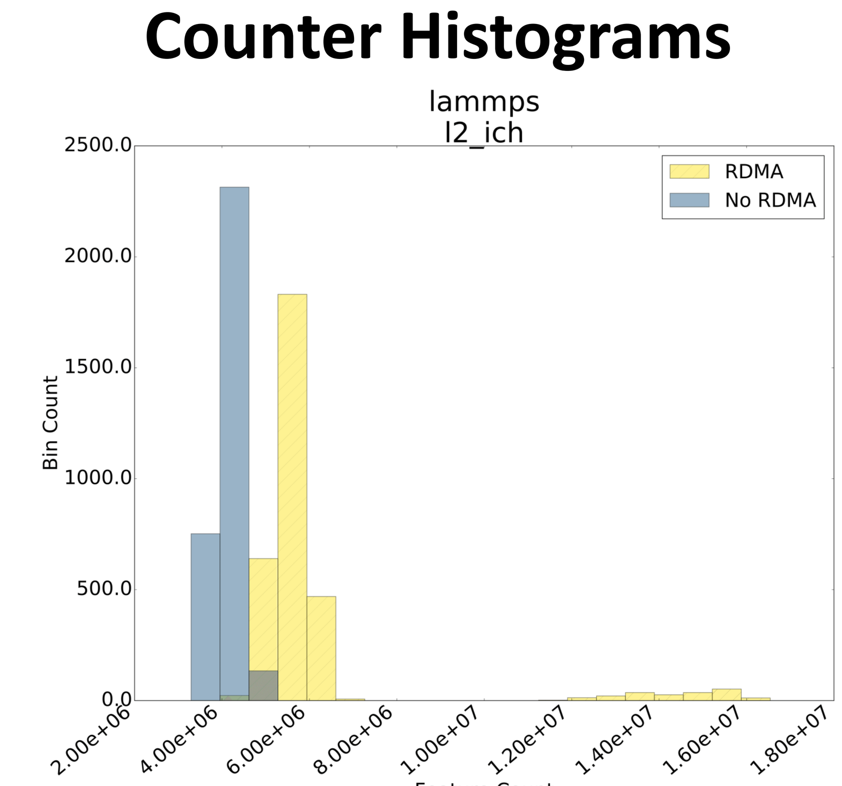
- Determining RDMA impact can be challenging:
  - Hard to determine traffic amount on target node
- NiMC affects each workload differently
- Machine learning correlates micro-performance to RDMA
  - Evaluated 17 performance counters
  - Importance of counter varies widely across workloads

Relevance of Counters



Machine learning shows five important L2 features for NiMC detection (LAMMPS)

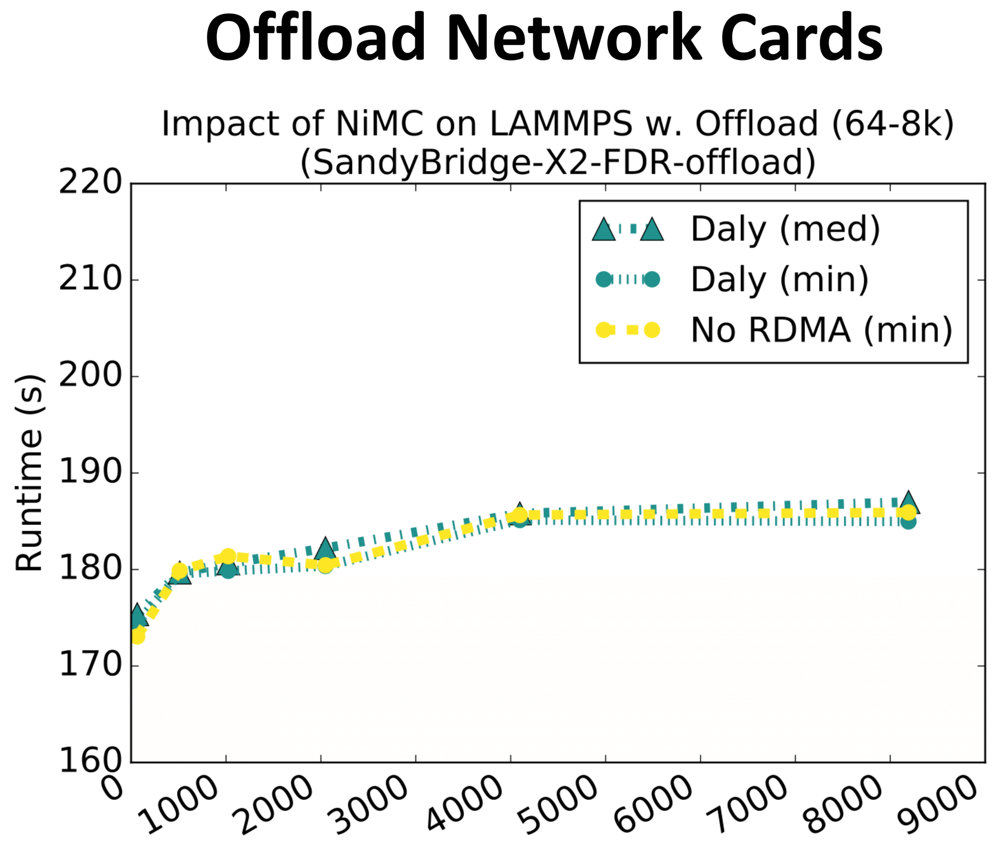
Counter Histograms



Histograms across 3,200 processes, show outlier processes on far right

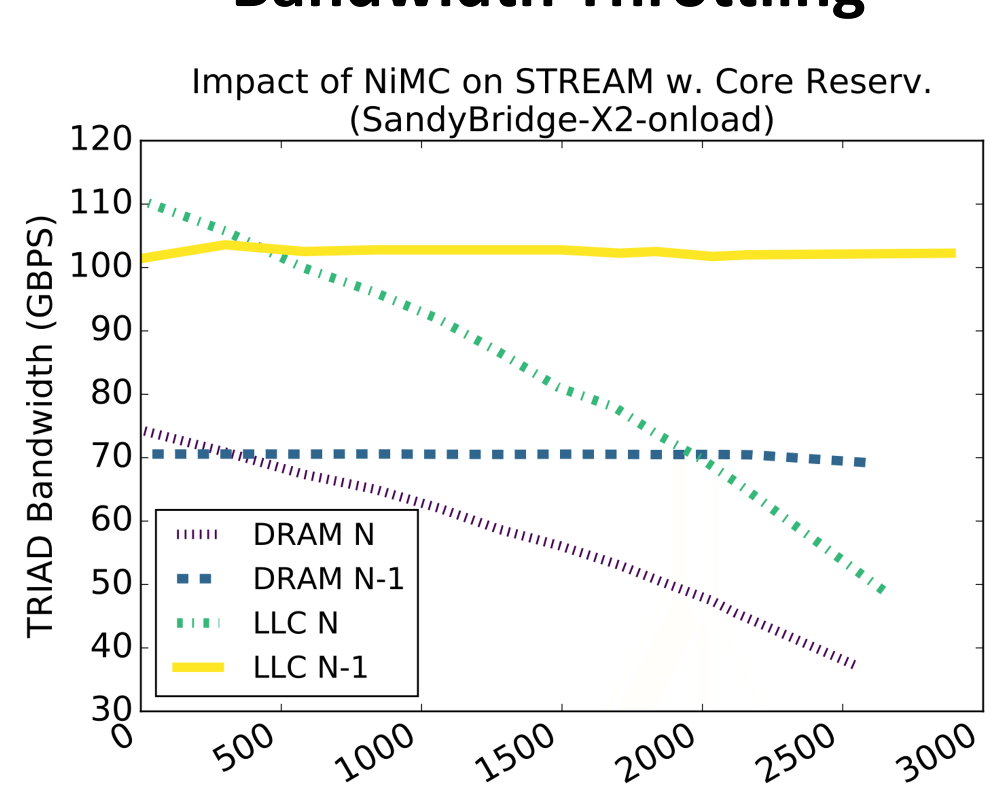
Mitigation Strategies

Offload Network Cards



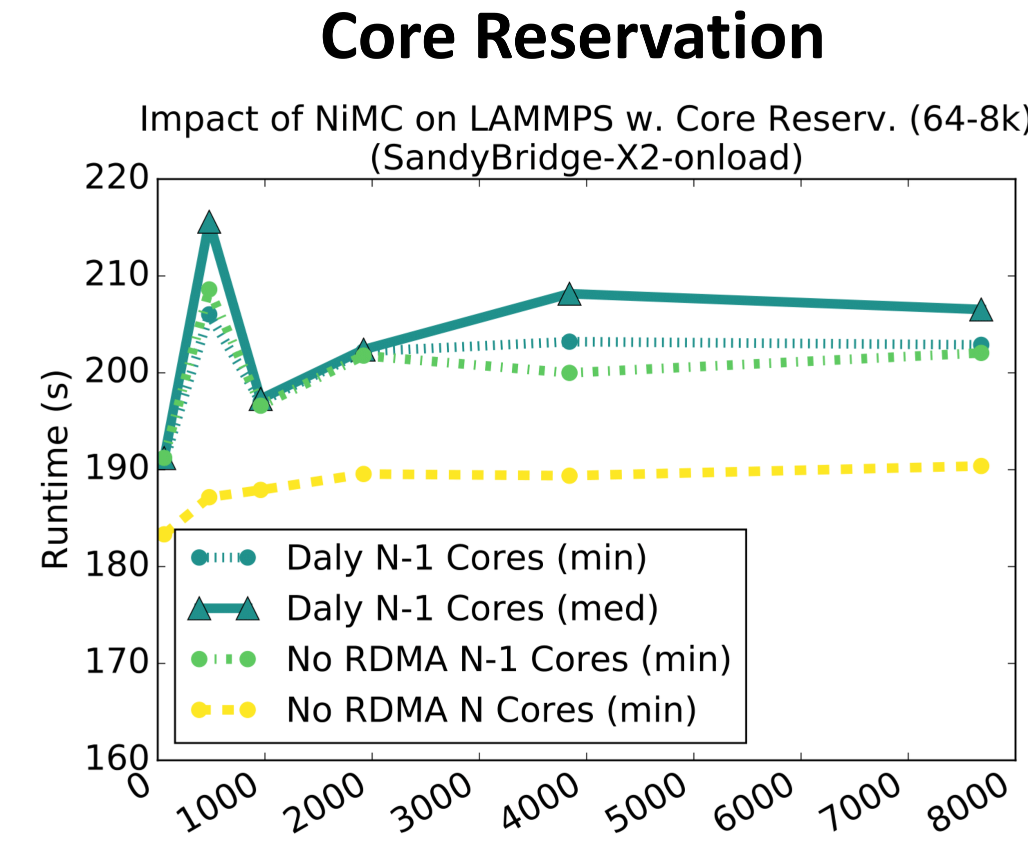
Offload cards can offer scalable solutions for modern CPUs at increased cost

Bandwidth Throttling



Throttling below (500MBps) provides benefits vs. core reservation

Core Reservation



Core reservation provides a scalable solution when throttling is not enough