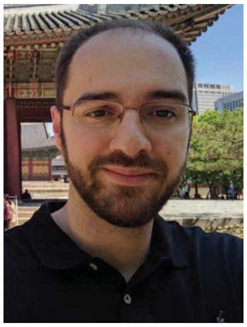
CNEC | Consortium for Nonproliferation Enabling Capabilities

# RAFFI YESSAYAN



University: North Carolina State University

Advisor: Yousry Azmy

Lab Mentor: Sebastian Schunert

Nuclear Science & Technology Directorate Nuclear Engineering Methods Development
Idaho National Laboratory

## Project Title

Transport Code Behavior with Latency Dominated Communication on Shared HPCs

## Project Objective

Quantify and identify ways to mitigate the severe latency effects incurred in high-utilization HPC systems.

## Project Description

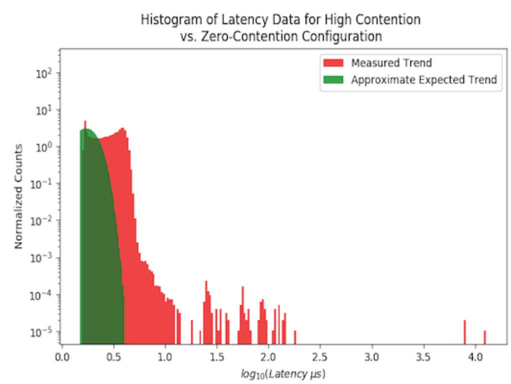
Previous work with both massively and moderately-sized parallel codes has shown unexpected scaling behavior which has detrimentally impacted parallel efficiency. Specifically, earlier implementations of the Parallel Block Jacobi and Parallel Gauss-Seidel iterative schemes for solving the radiation transport equation exhibit an unexplainable growth in the execution time per iteration with increasing number of processors. It is commonly thought that the cost of a communication operation on an HPC system is relatively constant with increasing number of participating processors and that the latency cost is miniscule. However, this knowledge contains the implicit assumption that the system is relatively un-congested and that the message sizes are relatively large. For modern, shared HPCs, high utilization (and consequently, high contention), are an everyday occurrence. Additionally, for massively parallel codes implementing spatial domain decomposition, the goal is to subdivide the work as finely as possible. This leads to myriad communications with small data sizes and relatively significant latency costs.

Previous analysis of the Falcon HPC system at INL to model latency had been performed during maintenance outages on a system with zero load by measuring the latency of single point-to-point messages. This produced results that showed a latency cost on the order of ~1.5us +/- 0.5us. However, high level analysis of the communication behavior of a massively parallel spatial domain decomposition radiation transport code indicated that the behavior measured during the outage was not representative of runtime behavior. Further analysis of the HPC, both under high load, and with more simultaneous communication was needed to quantify the difference and find ways to mitigate its effect on parallel efficiency.

## Project Relevance to Nuclear Nonproliferation

As deterministic transport problems have become increasingly high fidelity, the amount of computational resources required has dramatically increased. This led to the adoption of parallel and massively parallel methods. Better understanding of the increasing communication costs of these methods can drive significant efficiency gains and allow further increased fidelity.

## Products and Outcomes of Project




Working with both the HPC and methods development staff, further analysis of the communication latency was performed. This data showed latency variations several orders of magnitude higher than the no-load data and was used to generate a simulation model of a massively parallel transport iteration. This simulation indicates that the accumulation of high variance communication costs drives the parallel efficiency loss of the code. We are investigating mitigation strategies that may include subdomains that include a larger number of cells or lumping multiple small messages into fewer larger messages.

## Publications and Reports

R. YESSAYAN, S. SCHUNERT, Y. AZMY, "Behavior of Highly Communicating Codes Under Latency Dominated Regimes on High Utilization Shared HPCs," PHYSOR 2018, Cancun, Mexico, April 22-26, 2018, In Preparation.