Cover Page

NSF Faculty Early Career Development (CAREER) Program 2024
Directorate for Computer and Information Science and Engineering (CISE).

Project Title:

Toward a Robust Framework for Real-time Distributed Machine Learning in Dynamic and Zero Trust Environments using Anomaly Detection and Resilient Network Structures

Principal Investigator:

Dr. Duong Nguyen
Assistant Professor
Department of Electrical Engineering and Computer Science
University of Wyoming
307-766-2279
duong.nguyen@uwyo.edu

Amount requested: $500,000

Project Summary

Overview: Real-time distributed machine learning (RDML) has fueled the emergence of many applications in everyday life, from environmental sensing and traffic monitoring to self-driving cars and space exploration. A system using RDML consists of various hardware and software components to collect data and execute decentralized machine learning algorithms to formulate predictions, make decisions, and perform offline analysis. Such a system is typically controlled by one owner (e.g., a state, agency, or organization) and denoted as an autonomous intelligent group (AIG) in this proposal.

The deployment of an AIG often has limited coverage due to insufficient resources and other constraints (e.g., legal and political). To overcome this limitation, different AIGs connect and exchange data and information, leveraging diverse data input sources to boost the performance and accuracy of their RDML process. However, interconnecting AIGs faces two main challenges: (1) Resiliency: an AIG may receive incorrect data shared by other AIGs due to accidental errors or malicious intent and require mechanisms to prevent or minimize the damages caused by adversary events; and (2) Privacy/Security: an AIG may hesitate to share the information with other AIGs since such sharing could reveal sensitive details about the AIG internals and some collected data that its owner wants to keep protected. However, selfishness hurts collaboration and trust among AIGs.

Intellectual Merit: Although both Resiliency and Privacy/Security requirements are critical for any AIG, in this project, we focus on addressing the resiliency challenge. Specifically, our objectives are:

1. Analyze the impact of faults on the performance (convergence and accuracy) of RDML using static analysis enhanced with machine learning techniques. Results from this analysis will provide insights to develop effective techniques to classify and filter faulty information.
2. Use random matrix theory to analyze the structural and dynamical characterizations of AIG networks and their influence on the resilience of AIG to faults. The result will be used to design efficient and robust AIG structures/architectures.
3. Based on the results from the two former tasks, develop a holistic approach for designing a resilient RDML framework in a dynamic and zero-trust environment.

We note that to help an AIG protect its sensitive data while still honoring the commitment/contract of data sharing, differential privacy techniques (aggregation, masking, noise addition, randomization, etc.) are used to alter the data before releasing them. Such transformed data has the potential to be incorrectly flagged as faulty or disturb the convergence of RDML. We will consider this factor when developing the static and matrix analysis techniques. We will benchmark the developed techniques on a distributed computing cluster in the PI's lab.

Broader Impacts: The proposed project will identify primary features of faulty data exchanged among AIGs and provide insights into optimal AIG topological structure for tolerating adversary activities. These insights lead to novel algorithms and design principles for RDML systems, advancing research on robust and secured distributed machine learning for real-time systems. The proposed project will advance the PI early career through research that addresses a pressing computing research problem with a wide range of practical applications.

An important component of the proposal is promoting education excellence. The execution of the project involves and engages graduate and undergraduate students in research activities. Students will be introduced to the general framework of distributed computing, and be supervised to apply this framework to a distribute machine learning problem that he/she is most interested in. Results from this project will help the PI develop an advanced course in Distributed Optimization and Machine Learning for both graduate and senior undergraduate students.

Through University of Wyoming outreach initiatives (K-12, the Summer High School Institute program), we will introduce our research to a broader audience through lab demonstration and on-site hands-on projects for high school students and teachers.