Parallel AI and Systems for the Edge (PAISE)

Applications involving voluminous data but needing low-latency computation and local feedback require that the computing be performed as close to the data source as possible --- often at the interface to the physical world. Communication constraints and the need for privacy-preserving approaches also dictate the need for computing at the edge. Given the growth in such application scenarios and the recent advances in algorithms and techniques, machine learning and inference at the edge are unfolding and growing at a rapid pace. In support of these applications, a wide range of hardware (CPUs, GPUs, ASICs) is venturing farther away from the center, closer to the physical world. The resulting diversity in edge-computing hardware in terms of capabilities, architectures, and programming models poses several new challenges.

At the edge, several applications often need to be scheduled concurrently or serially. Some applications may need to be run continuously, a few in anticipation of certain events, whereas others may need to be run when particular events occur, causing a need to unload other applications and dedicate resources to them. Situations may also warrant running applications in sandboxes for privacy, security, and resource allocation reasons. A future with heterogeneous edge hardware and multiple applications sharing the hardware and energy resources is imminent.

Deploying and managing applications at the edge remotely, and building in multienancy to support applications with various resource constraints and runtime requirements, present a challenge that requires cooperation and coordination between the various components of the software stack. Mechanisms need to be devised that communicate both data and control with the applications in order to fine-tune their behavior and change the operational parameters. Coupling these edge applications with centrally located HPC resources and their applications, realizing the computing continuum, also opens up many research areas.

As we push more toward edge-enabled networks of devices, we inherit a setting where resources are deployed away from the safety of secure indoor spaces, often in the midst of a bustling urban canyon, and exposed to physical and cybersecurity threats. Deployed and interconnected predominantly over public networks, these systems have to be designed with cybersecurity as a first-class design citizen, rather than introduced as an afterthought.

The goal of this workshop is to gather the community working in three broad areas:

- processing — artificial intelligence, computer vision, machine learning;
- management — parallel and distributed programming models for resource-constrained and domain-specific hardware, containers, remote resource management, runtime-system design, and cybersecurity; and
- hardware — systems and devices conducive to use in resource-constrained (energy, space, etc.) applications.

The workshop will provide a critically needed opportunity to discuss the current trends and issues, to share visions, and to present solutions. This year's workshop will include a keynote, five papers, three invited talks and a panel on edge computing and inference at the edge.

DOI 10.1109/IPDPSW.2019.00143

Program Committee

- Prasanna Balaprakash, Argonne National Laboratory, USA
- Juan Pablo Bello, New York University, USA
- Cristiana Bentes, Universidade do Estado do Rio de Janeiro (UERJ), Brazil
- John Willian Branch Bedoya, Universidad Nacional de Colombia, Colombia
- Sergio Armando Gutiérrez Betancur, Universidad de Medellín, Colombia
- Kyle Bingman, Air Force Emerging Technologies Office, USAF, USA
- Charlie Catlett, Argonne National Laboratory, USA
- Peter Dinda, Northwestern University, USA
- Nicolas Erdody, Open Parallel, New Zealand
- Felipe M. G. França, Universidade Federal do Rio de Janeiro (UFRJ), Brazil
- Nicola Ferrier, University of Chicago, USA
- Dennis Gannon, Indiana University Bloomington, USA
- Eric Van Hensbergen, ARM Research, USA
- Christine Kendrick, City Of Portland, USA
- Sandip Kundu, University of Massachusetts Amherst, USA
- Priscila Machado Vieira Lima, Universidade Federal do Rio de Janeiro, Brazil
- Henrik Madsen, Technical University of Denmark, Denmark
- Leandro Marzulo, Google LLC, USA
- Alan Mainwaring, Intel Corporation, USA
- Eric Matson, Purdue University, USA
- Sanjay Padhi, Amazon Web Services, USA
- Vivien Rivera, Northwestern University, USA
- Koichi Shinoda, Tokyo Institute of Technology, Japan
- Michela Taufer, The University of Tennessee, Knoxville, USA
- German Sánchez Torres, Universidad del Magdalena, Colombia
- Jerry Trahan, Louisiana State University, USA
- Juan Rodrigo Sanz Uribe, National Coffee Research Center-Cenicafé, Colombia
- Sean Shahkarami, University of Chicago, USA
- Ramachandran Vaidyanathan, Louisiana State University, USA
- Kazutomo Yoshii, Argonne National Laboratory, USA

Organizing Committee

- Pete Beckman, Mathematics and Computer Science Division, Argonne National Laboratory, USA, E-mail: beckman@mcs.anl.gov
- Rajesh Sankaran, Mathematics and Computer Science Division, Argonne National Laboratory, USA, E-mail: rajesh@mcs.anl.gov