Knowledge Product: Intelligence at the Point of Need (for Tactical Edge Operations)

ARL POC: *Brian Panneton, brian.c.panneton.civ@army.mil, (410) 278-5345*
Consortium POC: *Prashant Shenoy, shenoy@umass.edu, (413) 577-0850*

OPPORTUNITY: Research continues to reduce the size, weight, and power requirements of lower-end (affordable) commodity hardware, making it attractive to deploy as 'smart-sensors' at the tactical edge. These hardware advances, however, result in increasingly specialized devices that are designed for a particular purpose, such as radios for communication and accelerometers to detect motion. The resource-constrained and purpose-driven nature of these devices introduces heterogeneity challenges that must be overcome to increase sensing versatility in the battlefield and afford intelligence at the point of need.

GOAL: Enable multi-purpose exploitation of distributed, heterogeneous, and multi-vantage commodity edge devices and sensors in complex and contested edge environments. Data from multiple modalities must be fused and the computational workload distributed across heterogeneous edge devices to provide real-time situational awareness and circumvent the resource constraints and adversarial denial that may occur at the tactical edge.

NEW SCIENCE: The IoBT CRA has advanced multi-purpose sensing/computing exploitation at the tactical edge through seven IoBT innovations: (1) LTE-based pervasive sensing, (2) rich semantic scene analysis by fusing ultra-wide-band radar with vision and LiDAR, (3) reconstruction of audio by frequency unfolding of accelerometer data, (4) model partitioning for distributed prediction, (5) sensor actuation and purpose driven prioritization of image areas, (6) uncertainty-minimizing actuated multi-sensor load assignment, and (7) robust methods for accelerating uncertainty quantification for deep neural networks at the edge.

The 4G LTE-based pervasive sensing is a form of unconventional, non-visual, non-line-of-sight sensing that is able to passively detect the presence and respiration rate of humans using commodity devices intended for communication. This first-of-its-kind research can be expanded to future generations of the RF spectrum to provide resiliency and opportunistic exploitation for situational awareness. Distributing partitioned neural network models across multiple resource constrained edge devices enables more flexibility and adaptability for inference. The ability to choose between processing data closer to the sensors or reach back for additional computing resources, if the network is not congested, increases flexibility in carrying on with the mission under a wide range of network conditions. The research into robust methods for accelerating uncertainty quantification for deep neural network models at the edge produced a benchmark called URSABench. URSABench characterizes accuracy-robustness-speed-storage trade-offs of deep learning models on embedded hardware, making them easier to compare and contrast.

SCIENTIFIC CHALLENGES AND RISK: The LTE based sensing capabilities have an assumption that the LTE infrastructure continues to be operational. If there are issues with the infrastructure, there is risk due to the dependence on beacons from the cell towers. Additionally, a challenge is dealing with weak signals from distant towers and high interference. Towers that are farther away provide less sensing capability. IoBT devices must overcome rapid event changes and still make accurate predictions and decisions in the presence of dynamic topologies, disconnected/intermittent/limited bandwidth and other environmental factors. A challenge for distributed prediction is to decide how to prioritize information processing when resources are being actively constrained. Distributed inference requires sending intermediate inference outputs over a wireless network, which can be detected and localized by adversaries based on received RF signatures, leading to the loss of the device.

RESULTS FROM EXPERIMENTATION: Unconventional sensing using commercial LoRa wireless radios can detect humans through five concrete walls, while yielding 2x increase in sensing range. Distributed partitioned neural networks have yielded a 200% reduction in detection latency over non-partitioned networks, while maintaining accuracy. All RF-sensing components have been tested in actual urban settings to demonstrate efficacy. Work enabling in-situ uncertainty estimation, as well as multimodal (acoustic and seismic) sensing and in-situ target classification analytics has also been tested on multimodal data collected from ARL's Robotics Research Collaboration Campus at Graces Quarters. It demonstrated target identification with high accuracy and improved calibration of neural network confidence. These analytics have also been tested on ARL's Acoustic-seismic Classification and Identification Data Set (ACIDS) with possible follow-up use in ARL's DARTS program and the Tailored AI at the Tactical Edge (TATE) capability pillar of the AI Strategic Challenge (AISC) under the Technical Cooperation Program (TTCP). Another line of work, CLIO, comprising a novel approach for neural network model partitioning for distributed prediction, has been run successfully on the Gap8 embedded Neural Accelerator platform. It performs better than JALAD, NeuroSurgeon, and 2-step pruning in most cases.

TECHNOLOGY TRANSITION: Multi-modal sensing innovations developed in the IoBT CRA are transitioned to MassAITC NIH Center on mobile/wearable sensing and AI, where an Alliance PI also serves on the Technology Identification and Training Core. These innovations also enrich the IoBT Digital Twin system being developed in an active transition to Boeing Inc. that encapsulates several IoBT innovations. Additionally, the research on understanding the tradeoffs between uncertainty, time, and resources consumed resulted in a benchmark framework called URSABench. An ARL PI and collaborator on URSABench has since accepted a position at the Joint AI Center (JAIC). We are investigating potential transition paths through the Joint AI Center.

IMPACT ON THE SCIENTIFIC COMMUNITY: The work produced multiple academic publications in premier peer-reviewed conferences and journals including ACM SenSys, ACM IMWUT, DCoSS, IEEE, ICML, ICCCN, UAI, USENIX Security, and WWW. The research has been recognized by best paper awards at RTSS and SenSys and the acceptance into two books, Statistical Inference for Engineers and Data Scientists (Cambridge U. Press) and Energy Management (World Scientific Press). Members of the research team were requested to be chairs for conferences at IEEE Infocom'21 and IoTDI'20 as well as an editor for the IEEE Open Journal of Signal Processing. Additional recognition was designated to researchers by admittance for Society Fellow Honors at IEEE, ACM and NAE.

POTENTIAL CAPABILITIES:

A key element in winning against a near-peer adversary is to induce flexibility by increasing available courses of action. Enabling multi-purpose exploitation of distributed, heterogeneous, and multi-vantage commodity edge devices and sensors in complex and contested edge environments improves flexibility and enhances understanding. A recent Focused Excursion organized by DEVCOM and FCC (a collaborative investigative process that supports concept and capability development with participation from FCC DoC, DoIS, DAC, and ARL) investigated IoBT implications on select learning demands of the Army Concepts Framework 2040. Among other hypotheses, the excursion suggested the hypothesis that "IoBT will increase situation awareness by exploiting ubiquitous, multimodal sensors, unconventional sensing, and edge computation in the operating environment". Current modernization efforts are working towards intelligence at the point of need. Better sensing capabilities empower better decision making. Distributing computation across heterogeneous edge nodes and making use of unconventional and multimodal sensing offers more flexibility to meet mission needs, while enhancing technologies at the edge and providing more effective situational awareness under adversarial conditions.

Experimental Demonstrations:

See videos on IoBT "AI at Point of Need" at: https://abdelzaher.cs.illinois.edu/RMB22-Demos.html

Key Publications*:

1. T. Abdelzaher, **A. Taliaferro**, P. Sullivan, **S. Russell**, "The Multidomain Effects Loop: From Future Concepts to Research Challenges," In Proc. *SPIE Defense and Commercial Sensing*, April 2020.
2. E. Blasch, **T. Pham**, C.Y. Chong, W. Koch, H. Leung, D. Braines, and T. Abdelzaher, "Machine Learning/Artificial Intelligence for Sensor Data Fusion–Opportunities and Challenges," *IEEE Aerospace and Electronic Systems Magazine*, *36* (7), pp. 80-93, 2021
3. X. Chen, D. Ganesan, J. Gummeson, and M. Rostami, "COCOON - A Conductive Substrate-based Coupled Oscillator Network for Wireless Communication," *ACM SenSys* 2021.
4. A.D. Cobb, **B. Jalaian**, **N. Bastian**, and **S. Russell**, "Robust Decision-making in the Internet of Battlefield Things using Bayesian Neural Networks," In Proc. *Winter Simulation Conference (WSC)*, 2021.
5. S. Fang, J. Huang, C. Samplawski, D. Ganesan, B. Marlin, T. Abdelzaher, and **M. Wigness**, "Optimizing Intelligent Edge-clouds with Partitioning, Compression and Speculative Inference," In Proc. *IEEE Military Communications Conference (MILCOM)* (pp. 892-896), 2021.
6. J.Z. Hare, C.A. Uribe, **L. Kaplan**, and A. Jadbabaie, "Toward Active Sequential Hypothesis Testing with Uncertain Models," In Proc. *60th IEEE Conference on Decision and Control (CDC)* (pp. 3709-3716), 2021.
7. Y. Hu, S. Liu, T. Abdelzaher, **M. Wigness**, and **P. David**, "On Exploring Image Resizing for Optimizing Criticality-based Machine Perception, In Proc. *IEEE 27th International Conference on Embedded and Real-Time Computing Systems and Applications (RTCSA)* (pp. 169-178), 2021.
8. J. Huang, C. Samplawski, D. Ganesan, B. Marlin, and **H. Kwon**, "CLIO: Enabling Automatic Compilation of Deep Learning Pipelines Across IoT and Cloud," In Proc. *26th Annual International Conference on Mobile Computing and Networking*, 2020.
9. D. Liu, **P. Wang**, T. Wang, T. Abdelzaher, "Self-contrastive Learning Based on Semi-Supervised Radio Modulation Classification," In Proc. *MILCOM*, December 2021
10. C. Samplawski and B. Marlin, "Towards Transformer-Based Real-Time Object Detection at the Edge: A Benchmarking Study," In *IEEE Military Communications Conference (MILCOM)* (pp. 898-903), 2021.
11. A. Singh, L. Noor, and M. Srivastava, "I Always Feel Like Somebody's Sensing Me! A Framework to Detect, Identify, and Localize Clandestine Wireless Sensors," In Proc. *USENIX Security*, 2021.
12. M. Vadera, A. Cobb, **B. Jalaian**, and B. Marlin, "URSABench: Comprehensive Benchmarking of Approximate Bayesian Inference Methods for Deep Neural Networks, In Proc. *37th International Conference on Machine Learning*, 2020.
13. T. Wang, S. Yao, S. Liu, J. Li, D. Liu, H. Shao, R. Wang, T. Abdelzaher, "Audio Keyword Reconstruction from On-Device Motion Sensor Signals via Neural Frequency Unfolding," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT)*, Volume 5, Issue 3, September 2021.
14. Z. Wang, A. Singh, L. Garcia, M. Srivastava, "UWHear: Through-wall Extraction and Separation of Audio Vibrations Using Wireless Signals," In Proc. *18th Conference on Embedded Networked Sensor Systems (SenSys)*, November 2020.
15. F. Yuda, X. Yaxiong, D. Ganesan, and X. Jie, "LTE-based Pervasive Sensing Across Indoor and Outdoor," *ACM SenSys*. 2021.

*Note: Names in **blue** are government co-authors.