WASP: A Wearable Super-Computing Platform for Distributed Intelligence in Multi-Agent Systems

Chinmaya Patnayak
Department of Electrical and Computer Engineering
Virginia Tech
Blacksburg, Virginia, USA
Email: cpatnayak@vt.edu

James E. McClure
Advanced Research Computing
Virginia Tech
Blacksburg, Virginia, USA
Email: mcclurej@vt.edu

Ryan K. Williams
Department of Electrical and Computer Engineering
Virginia Tech
Blacksburg, Virginia, USA
Email: rywilli1@vt.edu

2021 IEEE High Performance Extreme Computing Conference (HPEC) | 978-1-6654-2369-4/21/$31.00 ©2021 IEEE | DOI: 10.1109/HPEC49654.2021.9622784

***Abstract*—Autonomous unmanned aerial vehicle (UAV) systems have broad applications in surveillance, disaster management, and search and rescue (SAR) operations. Field deployments of intelligent multi-UAV systems are heavily constrained by available power and networking capabilities, and limited computational processing resources which are needed to reduce large volumes of on-board sensor data in real-time. In this work, we design a *WearAble Super-Computing Platform (WASP)* to address such challenges associated with multi-UAV deployments in remote field environments based on a human-in-the-loop (HITL) design. The WASP system is an advanced edge computing instrument designed from commodity embedded processing devices interconnected through an on-board Ethernet network. Networking is further extended through wireless networking capabilities to communicate with UAVs. Computational workloads and storage are orchestrated as discrete containers across WASP and the UAVs, which accounts for processor heterogeneity and time-varying workloads that must adapt dynamically to account for unpredictable failures of wireless networking in the field. We use our prototype to demonstrate advantages in terms of power management, redundancy, robustness, and human-robot collaboration in challenging field environments.**

***Index Terms*—Artificial Intelligence, Inference, Distributed Systems, Cyber physical Systems, Edge Computing, Wearable Computing**

I. Introduction

Autonomous robots have significantly extended human capabilities for tasks ranging from nano-robots for medical diagnosis to rovers for space exploration [1], [2]. Unmanned aerial vehicles (UAVs) in particular have been of significant interest due to their potential impact in diverse fields such as transportation, surveillance, defense, agriculture, search and rescue (SAR), and photography [3]–[8]. However, for autonomous UAVs (and other agents) to be viable in field applications, they must overcome critical practical limitations. Specifically, available on-board power and computational resources are key constraints for autonomous agents, especially UAVs. With evolving sensors and their ubiquity, management and concerted aggregation of vast streams of localized data are essential to meet cognitive demands from the system. Connectivity and management of big data have emerged as integral challenges that must be addressed at the system design level to achieve scalable deployments.

Advancements in machine intelligence are driving shifts in traditionally human-driven applications through increasing reliance on autonomy and robotic agents. Modern Search and Rescue (SAR) stands out as an emerging area that requires complex interactions and information flow among multiple agents in a system including human responders, computers, and robots. The introduction of computing and robotic agents have not only helped overcome several limitations held by human-only missions including inaccessibility, situational awareness, viability, and scalability, but have also enhanced safety of the human responders in the process [9]–[11]. As of today, advanced ground station equipment and intelligent mobile agents have successfully been deployed in real situations otherwise prohibitive for human-only operations [11]–[18]. With data generation rates continuing to outpace growth in processing capability, solutions that leverage intelligent systems will continue to suffer in terms of collaboration, viability, and scalability.

In this work, we propose a new multi-agent system for SAR operations by prototyping a wearable and mobile compute cluster to overcome practical challenges that limit the capabilities of such a system. The primary objectives behind developing the mobile compute cluster are:

- *Coordination*: Enhanced coordination among human responders and robotic agents in the system by means of multi-modal data aggregation, information sharing, and situation-aware fallback mechanisms;
- *Computing*: A high-performance accelerated surface for compute-intensive workloads and real-time intelligence that is scalable and robust for field applications;
- *Distribution*: Offloading computational and physical payloads from field (aerial) agents to reduce on-board power footprint.

The following section describes the evolution of multi-agent systems in SAR and impact of modern computing paradigms such as distributed and edge computing. Section III defines our proposed system, including the WASP and interaction among agents in the system. In Section IV, we quantitatively evaluate the computational characteristics for our design prototype and demonstrate its advantages over existing systems.

II. BACKGROUND

Introduction of unmanned agents, especially UAVs, in SAR stems from successful deployments in military applications. The most significant constraints for UAVs and similar field autonomous agents [11], [19] can be summarised as:

- Insufficient on-board power, seldom supplied by a robust and continuous power source, to support viable flight times, sensors, and other payload;
- Insufficient on-board computational resources for operations at the edge such a machine learning inference for detection, path planning, and storage;
- Lack of continuously available low-latency, high-bandwidth channels for communication with other agents and/or with cloud based servers.

Power efficiency is critical to developing autonomous systems that can surpass human capabilities in a field setting; biological systems are remarkable in terms of their energy efficiency. A direct impact of insufficient on-board power for UAVs is short flight times ranging from 15 - 40 minutes which severely limits their operational viability. Physical payload further loads UAV motors and on-board computing based on the relationship depicted in Figure 1. The development of high capacity portable power sources continues to remain a work in progress. Investigation of power sources is out of scope for this work. The primary goals of our effort are to optimize computational systems within the context of the limitations associated with payload, computational resources, communication channels, and the interplay among multi-agent systems.

Autonomy in a SAR-like system may ultimately become a function of computational resources performing high-accuracy inferences. Collaborative action and coordination rely on networking capabilities and management. Within such a system, different kinds of computing elements may be made available, and it is desirable to map tasks to the particular processing element where they can be performed most efficiently. Factors influencing such a mapping include algorithmic constraints, data locality, and field power constraints. Software systems should be highly fault-tolerant and opportunistic in nature, particularly with respect to intermittent failures of wireless networking. Computational systems must be designed with these factors in mind. The advancement and maturation of such systems represent a basic computational barrier to the deployment of UAV teams that can operate autonomously in challenging field environments.

Adjusting for the saturation of Moore's law, over the last two decades we have continued to observe increasing computational power per unit die area, while the power consumption has continued to diminish. New workloads and efficient tools have brought about a shift away from general purpose computing and have renewed interest in hardware and software for specialized logic and operations in the form of FPGAs, GPUs, and ASICs. This recent paradigm shift has been further propelled by the advancement of machine learning techniques and increasing data volumes that benefit from specialized processing capabilities.

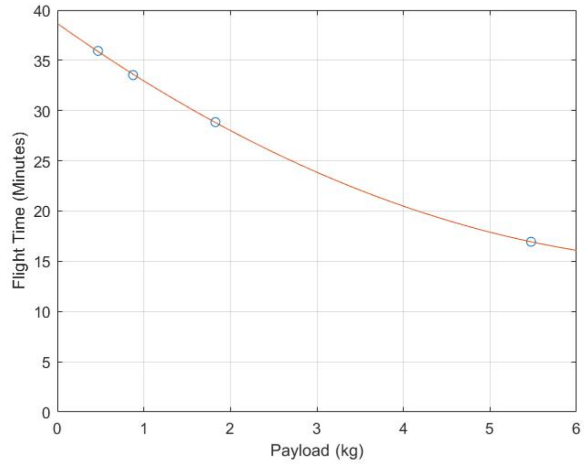

Fig. 1. Augmenting UAV with processing capabilities can reduce flight time due to increased payload. UAV carrying onboard processors must account for the weight of these elements in addition to their own power consumption.

A promising theme in overcoming computational limitations has been to offload tasks away from resource-constrained agents. Offloading computational processes from mobile agents allows for significant energy reduction. Previous work comparing on-board and offloaded computation on UAVs for video processing records more than 100x reduction in energy consumption [20]. The ability to offload computation also ensures that the system can scale as per changing situational requirements. Two common surfaces for offloading computation are cloud services and secondary local devices. Cloud services such as Amazon Web Services (AWS), Google Cloud Platform (GCP), and Microsoft Azure Services have evolved and matured over the last decade to allow multiple computational backends suitable for a particular task combined with a suite of supporting tools and resources. They have also made it possible to provision virtually infinite resources for applications. Despite the advantages and convenience, utilization of such services assumes a continuous high bandwidth and low latency channel to the internet. This restricts the utilization of such services for field applications such as SAR and disaster relief, which are motivating applications for our work [21]. Fog computing and edge computing [22], [23], in contrast to cloud computing, develops the idea of bringing computational resources closer to the source of data or sensors, by means of standalone devices available locally or embedded at the source of data. Bringing computation nearer to the source of data enables low latency results for time critical and remote applications. Ubiquity and physical form factors of such computing resources have also helped strengthen the scalability and viability of edge-based solutions. This work focuses on leveraging edge devices to address computational shortcomings for systems deployed in remote and harsh environments.

978-1-6654-2369-4/21/$31.00 ©2021 IEEE

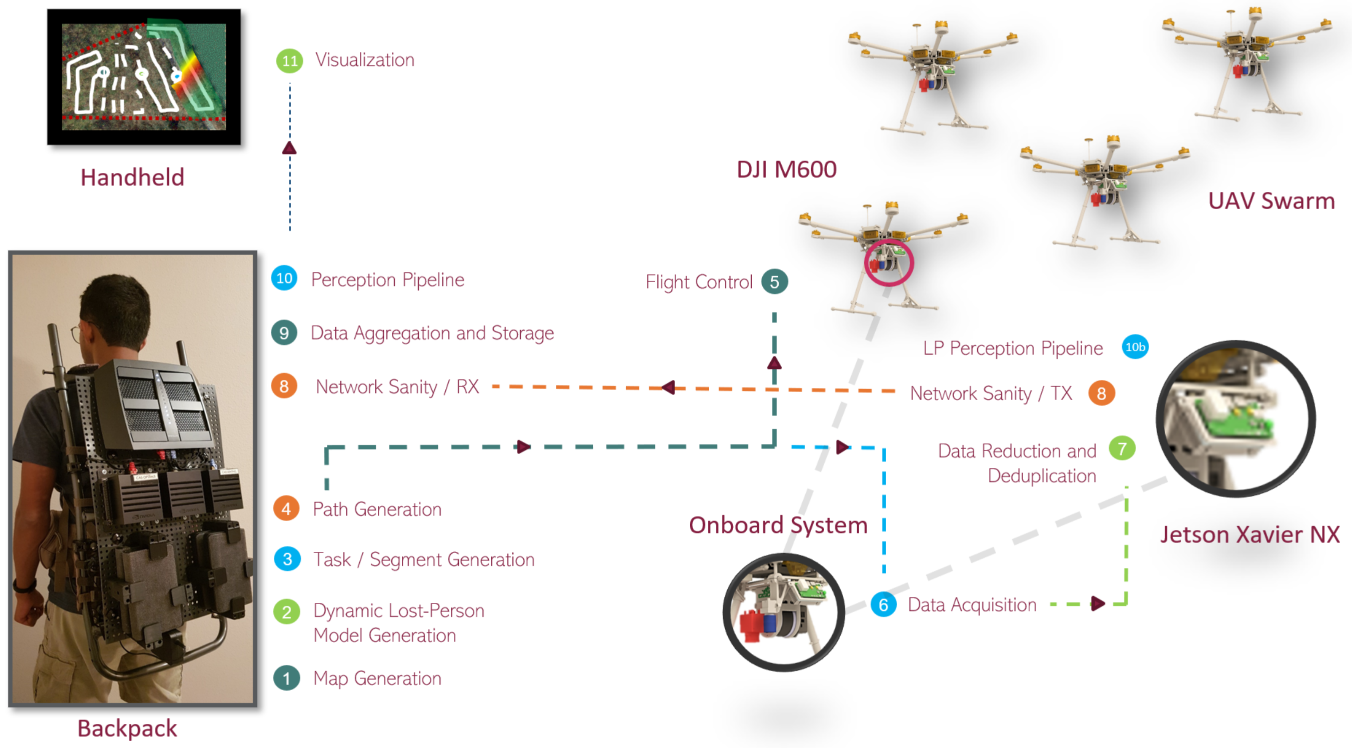

Fig. 2. **System Design and Interactions.** Edge computing approach to provide processing capabilities to a team of UAV. Computations can be offloaded to balance tradeoff between data locality and payload / flight time. Hardware elements are mounted onto a backpack frame that can be carried into the field. Compute resources within a cluster may be contained within the backpack or attached to UAV. Both network and compute resources may be heterogeneous.

Previous work in partitioning and offloading computation to edge devices has found applicability for many use cases especially surveillance. Significant work has been done to highlight the advantages of partitioning and offloading computation with edge devices in SAR-like applications, and devising optimal methods or triggers for the same [24], [25]. More prominent among such strategies were based on geographical proximity [21] and minimizing the net energy consumption of the system [26]. MILP techniques have also been applied and evaluated to extend distribution based on minimizing the net energy consumption of the system [27].

Another significant aspect of such field distributed systems is intelligent task prioritization and scheduling. We would consider this as a secondary level of power optimization by managing decisions and computations based on the availability of resources and context of computation. This eliminates redundancies and extraneous computational cycles that offer minimal rewards. Developments in heterogeneous task scheduling and real-time systems have been much more recent due to the increased availability of a wide range of devices and inter-compatibility owing to software standardization and community-led developments. Mobility of agents add to scheduling complexity. A novel Mobile Resource Aware (MRA) scheduling algorithm, proposed by Wan et al., extends generic stable scheduling algorithms to dynamic clients and servers [28].

Modern software workloads, such as Deep Learning Inference and highly parallelized path planning have also opened up unique challenges in terms of execution and scheduling. Pruned-down networks, at the cost of slight accuracy losses, have made it possible to run extensive deep learning solutions such as object detection and semantic segmentation on edge nodes providing real-time results [29]. Due to their inherently hierarchical nature, scheduling for deep learning inference might require a different approach depending on their network design. Such considerations may be particularly important in designing systems that are able to learn on the fly [23], [24], [30].

III. WASP Design and Development

A high-level design of the proposed system with its agents, their function and interactions are depicted in Figure 2. Decentralization in terms of functions, scalability, systemic redundancy and enhanced on-board capabilities were key motivational objectives driving our high-level design. The three major agents in our proposed system can be identified as:

- **Human Responder(s)**
  In contrast to traditional SAR operations, where the human responders are committed to the complete management and execution of the operation, our proposed system emphasizes on decentralization of core functions. The human responder(s) maintains responsibility for only critical and high-level decision making and control of a mission, incorporating tasks such as prioritization and

978-1-6654-2369-4/21/$31.00 ©2021 IEEE

observation. Aspects of the mission such as generating the body of knowledge, coordination, efficient path planning, device monitoring and detection are off-loaded to other computing agents in the system and are largely automated. The human interfaces with the other agents using a Graphical User Interface (GUI) exposed through a handheld device to manage and monitor the agents in the system. The degree of human intervention is expected to decrease with increasing autonomy in unmanned agents in the system and computational prowess.

- **Unmanned Agents (UAVs, UGVs)**
  Robotic agents supplant ground human responders responsible for the search and act as eyes in the ground (UGVs) and the sky (UAVs). Robots can keep the human responders away from risk-prone search areas, and augment understanding by offering alternate points of view (POVs). Significant improvements in UAV design and available power have enabled modern drones to host an array of sensors and lightweight processors. Payloads are much less restrictive for unmanned ground vehicles. Sensors used for SAR are expected to be cameras working in the visible and infra-red wavelengths and LiDAR. The on-board computer is responsible for data aggregation and transmission, and for running light-weight perception algorithms and/or fallback applications.
- **Wearable Mobile Compute Cluster**
  The standalone wearable compute cluster, which defines the most significant contribution of this work, enables us to address decentralization, systemic redundancies in case of failures, efficient data aggregation, enhanced coordination and understanding.
  The cluster bears the responsibility for heavy-lifting the central software stack that comprises mapping, task recommendation, prioritization, waypoint generation, path planning, intelligence software backbone, networking pipeline and the user interface. Additionally, the cluster supplements the mobile agents to meet the computational demands of the mission by limiting the latter's utilization of energy for tasks relating to mobility and information capture. Multiple heterogeneous computing surfaces contained in the cluster and efficient node management by software tools prevent loss of information. Such a tool also allows efficient flow and distribution of data and task execution to maximize platform utilization to generate high systemic throughput. Seamless data transactions across the agents and the backpack is realised by networking hardware on-board the compute cluster providing high bandwidth channels through Wi-Fi and Ethernet.

The design and prototyping of WASP were motivated by four pillar design decisions viz. Scalability, Portability, Safety and Ergonomics. The form factor of the compute cluster is derived from a hiking utility backpack, a common part of gear for rescue personnel. The backpack is designed to be standalone in terms of power, compute and networking. The cluster hosts three NVIDIA Jetson Xavier AGX to support compute loads and storage. Low power requirements, known efficacy in managing highly parallelized workloads including Deep Learning inference, thorough documentation and a mature SDK drove our choice of embedded hardware platform. The three Jetson Xavier AGX were interconnected through Ethernet on a Local Area Network managed through a Netgear Tri-Band Router. Networking was established between the cluster and smaller embedded computers (NVIDIA Jetson Xavier NX) mounted on UAVs through Wi-Fi. All components mounted on the WASP are powered through an array of four portable Li-Ion batteries each offering capacity of 85W through DC (5V, 9V) and AC (120V).

The physical construction of the WASP is achieved around an lightweight Aluminium frame from a hiking backpack with ergonomic straps and supports. A perforated PVC sheet, for its non-conducting and interfering properties, is used as a base for the embedded computers, networking router and batteries. All the components are attached together by off-the-shelf clamps or custom 3D printed connectors. The first finished iteration of the backpack holds 3 compute devices (maximum of 6) and 4 batteries on-board (maximum of 6).

Our computational design is intended to support scalable Human-in-the-loop (HITL) interactions. In typical SAR applications, UAVs would be deployed in conjunction with human search terms, and the HITL design is important to manage interactions between human searchers and UAV teams. Within this context, the WASP system must perform data reduction tasks so that information collected by UAV can be more easily digested by humans. The human responder in our HITL system maintains the high level control for defining the goals and perimeter of the mission. High definition GIS map along with victim parameters are fed to an implementation of dynamic lost-person behavior model [31]. On the basis of the behavior-model, the high definition map is broken down into sectors or tasks. Task prioritization could either be made manually by rescue personnel or recommended to the responder by a region and risk aware recommendation application [32]–[34]. On determining optimal traversal of task sectors and priority, a risk-aware path planning algorithm [35]–[37] generates paths for the unmanned agents in the form of navigation waypoints. The waypoints are then passed onto the UAVs and control signals are set.

The unmanned agents utilize the waypoints to follow a defined path. During this process, the agents collect and reduce data through compression and eliminating data redundancies while limiting the loss of information. The data is then transmitted back to the WASP, where the information is stored in a Distributed File System (DFS). Data on the cluster is managed through Hadoop Distributed File System (HDFS) across the nodes. The data is then acted upon by multiple instances of Containers running DL inference applications. Docker was utilized to develop custom containers to enable the isolation of such instances. Kubernetes was the choice of container orchestrator to scale the instances across the nodes on the WASP as well as the UAVs. The applications generate

978-1-6654-2369-4/21/$31.00 ©2021 IEEE

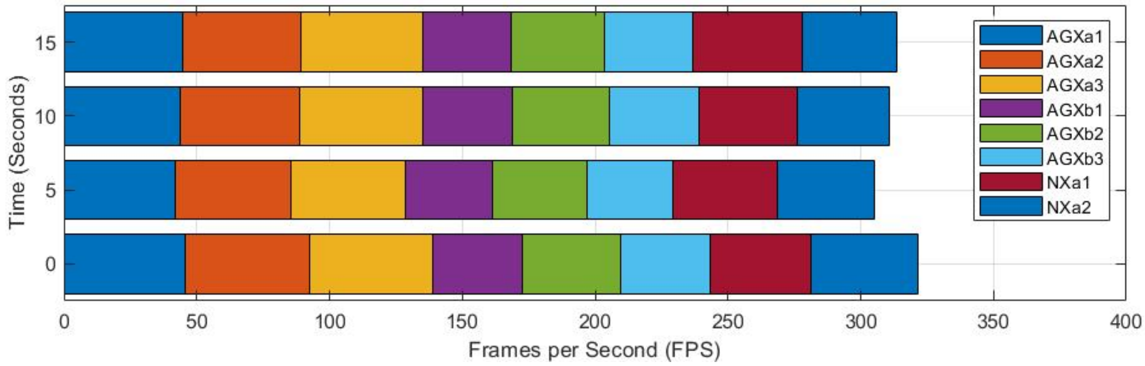

Fig. 3. Workload (Deep Learning inference, ResNet-Inception-V2) distribution for 8 instances of network running across 3 Nodes (2 Xavier AGX, 1 Xavier NX) at time t'=0, 5, 10, 15 sec. Each colored block represents average throughput recorded for each pod (container instance) until time t'.

the necessary analytics and visualizations to aids the human responder in making operational decisions. The WASP also extends a feedback mechanism by exposing a GUI to a human responder through a handheld device.

IV. Evaluation

This section evaluates the impact attributed to the addition of the wearable compute cluster in a SAR-type mission.

A. Computing

A complex system like the one we propose can be expected to run a fairly diverse set of software applications and complex algorithms for tasks including but not limited to image preprocessing, data reduction, computer vision, path planning and user interface. A large subset of such algorithms benefits from hardware and software parallelism and application-specific hardware. By utilising accelerated hardware platforms and efficient data and workload distribution infrastructure we extend the computational capabilities of traditional systems in use for SAR-like applications. We evaluated our proposed system based on real-time performance of workloads and the impact of distribution across multiple nodes in the system. For evaluation, we identified Deep Learning Inference as one of the most compute-intensive workloads for tasks (object detection, semantic segmentation, etc.) in SAR-like applications.

Single node inference throughput, measured in frames processed per second, was collected for compute nodes on the WASP (Xavier AGX) and UAV (Xavier NX) for four representative Deep Neural Networks including semantic segmentation [38], pose estimation [39], high-accuracy [40] and light-weight [29] object detection in Figure 4. We observed that the throughput across any given node for representative tasks was higher than the frequency of the fastest on-board sensor (RGB Camera, 60 FPS) thereby implying par real-time single node performance.

A more involved evaluation was made on the distribution of computational payload to improve systemic throughput. Distribution and collaboration ensure that the nodes are efficiently utilized and knowledge from various nodes is aggregated to develop a global body of knowledge for SAR-like applications. We measure distribution by means of increase in inference throughput for a single representative task over multiple computational surfaces. We revert to model parallelism wherein same copy of the model is executed in multiple instances (or Kubernetes pods).

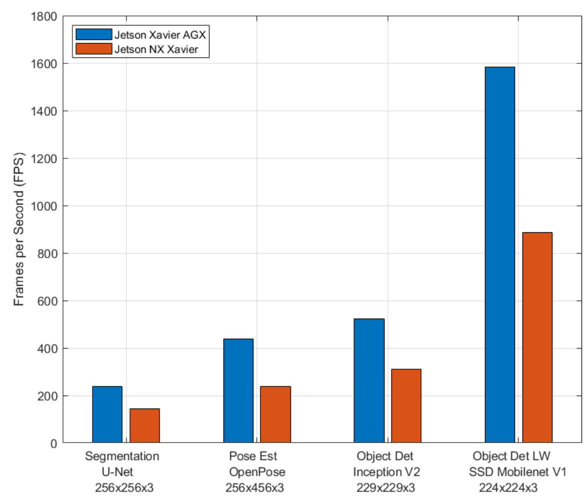

Fig. 4. Benchmarking results for video processing inference tasks on the WASP and UAV. The nodes in the system were observed to be sufficient to keep-up with the frame rate associated with generic cameras and UAV sensors.

To evaluate throughput in distributed inference, we provision two nodes on the backpack and a node mounted on a UAV, interconnected through the router through wired and wireless channels, respectively. We spawn a sample 8 Kubernetes pods running deep learning inference with ResNet-Inception-V2 network [40]) across those devices. The observations are summarized in Figure 3. We observe a consistent 300+ FPS throughput consistently distributed across nodes on devices which was sufficient for as many as 5 simultaneous camera feeds at 60 FPS.

978-1-6654-2369-4/21/$31.00 ©2021 IEEE

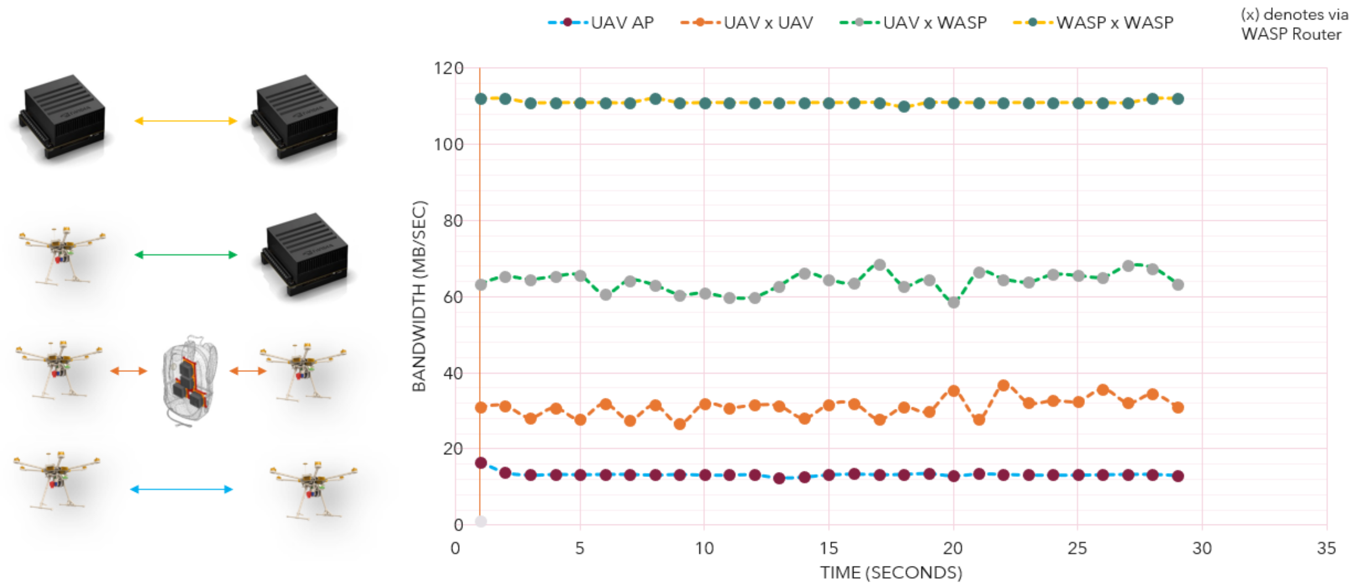

Fig. 5. Evaluation of network bandwidth for data transfers across agents (Top to Bottom: AGX to AGX via Ethernet, UAV to AGX via Wi-Fi, UAV to UAV via Wi-Fi through router, UAV-UAV through standalone access points). Data transfer rates are sufficient to offload video data for the small UAV teams considered in this work.

	Load	Type	Uptime
Session 1	Jetson AGX (30W) via USB-C	Idle	10h: 27m : 18s
Session 2	Jetson AGX (30W) via USB-C	DetectNet Object Detection (SSD / Mobilenet-v2)	06h: 52m : 54s
Session 3	Netgear Nighthawk XS6 AC4000 via AC	1 Hz ping across devices (3 x Wi-Fi; 3 x LAN)	05h: 15m : 23s

TABLE I
POWER BENCHMARKING: BATTERY LONGEVITY VS PAYLOAD

B. Connectivity

The ability to transmit large amounts of data such as video streams and databases with high bandwidth and reliable channels is essential for developing collaborative multi-agent systems. To assess the impact of the WASP with powerful on-board networking infrastructure we identified bandwidth as a representative parameter. The evaluation was made with iperf3 utility for Linux and bidirectional tests were conducted across all communication channels i.e. WASP-WASP (Ethernet), WASP-UAV (Wi-Fi), UAV-UAV (Wi-Fi) for a period of 30 seconds each.

We observed that introducing the router between the UAVs not only extended the communication range but also doubled the throughput for inter-UAV communication. We observed a 4x and 8x increase in bandwidth for transmission of the UAV with the backpack and inter-node communication of the backpack, respectively. This represents significantly faster data movement and processing when done on the backpack with respect on-board computing as is summarized in Figure 5.

C. Power Management

To evaluate the longevity of the WASP for field applicability, we subject the power source to three experimental loads. The loads includes an idle WASP node, a WASP node executing DNN-based object detection with the hardware configured on its highest power mode, and finally, the networking router connected to 3 wired and 3 wireless devices with a 1 Hz ping established across all devices and disabled power saving settings. The results are recorded in Table I. We determine our bottleneck at the networking infrastructure with an uptime of 315 minutes, which is more than 7 times the upper limit of the UAV flight times in our experiments when used as standalone agents in SAR missions.

V. CONCLUSION

By provisioning a functional prototype of a wearable compute cluster and software infrastructure, we demonstrate the impact and advantages of distributed intelligence for multi agent cyber-physical systems. This work extends existing systems in use in SAR by enabling scalability, efficiency in computational processes, robust collaboration and communication across agents by means of a new field agent. Our success with preliminary physical field-testing and evaluation of the wearable backpack motivates us to lay down future work including, but not limited to, development and optimization of task-specific payloads, intensive field testing and redesign into a more robust and rugged form factor.

978-1-6654-2369-4/21/$31.00 ©2021 IEEE

Acknowledgment

This research was supported by the National Science Foundation (NSF) National Robotics Initiative (NRI) under award 1830414. The authors would like to acknowledge Larkin Heintzman, Nathan Lau, Nicole Abaid for useful discussions and support pertaining to the computational requirements and design of the described system.

References

[1] J. Li, B. Esteban-Fernández de Ávila, W. Gao, L. Zhang, and J. Wang, "Micro/Nanorobots for biomedicine: Delivery, surgery, sensing, and detoxification," *Sci Robot*, vol. 2, no. 4, Mar. 2017.

[2] F. Zhou, R. E. Arvidson, K. Bennett, B. Trease, R. Lindemann, P. Bellutta, K. Iagnemma, and C. Senatore, "Simulations of mars rover traverses," *J. field robot.*, vol. 31, no. 1, pp. 141–160, Jan. 2014.

[3] M. Kamali, L. A. Dennis, O. McAree, M. Fisher, and S. M. Veres, "Formal verification of autonomous vehicle platooning," *Science of Computer Programming*, vol. 148, pp. 88–106, Nov. 2017.

[4] J. Liu and R. K. Williams, "Coupled temporal and spatial environment monitoring for multi-agent teams in precision farming," *Proc. IEEE Conf. Decis. Control*, 2020.

[5] P. Mukherjee, M. Santilli, A. Gasparri, and R. K. Williams, "Experimental validation of stable coordination for multi-robot systems with limited fields of view using a portable multi-robot testbed," *ISM*, 2019.

[6] J. Liu and R. K. Williams, "Optimal intermittent deployment and sensor selection for environmental sensing with multi-robot teams," *2018 IEEE International Conference on*, 2018.

[7] ——, "Monitoring over the long term: Intermittent deployment and sensing strategies for multi-robot teams," *2020 IEEE International Conference on*, 2020.

[8] R. Williams, "Collaborative multi-robot multi-human teams in search and rescue," in *Proceedings of the International ISCRAM Conference*, vol. 17, 2020.

[9] K. Iserson, "Injuries to search and rescue volunteers. a 30-year experience," *The Western journal of medicine*, vol. 151, pp. 352–3, 10 1989.

[10] I. Nourbakhsh, K. Sycara, M. Koes, M. Yong, M. Lewis, and S. Burion, "Human-robot teaming for search and rescue," *IEEE Pervasive Computing*, vol. 4, no. 1, pp. 72–79, 2005.

[11] M. Chalaris, G. Pallis, G. Statheropoulos, I. Balta, K. Mikedi, P. J. Vamvakari, A. Pappa, G. Xanthopoulos, M. Statheropoulos, S. Karma, and E. Zorba, "Use of unmanned vehicles in search and rescue operations in forest fires: Advantages and limitations observed in a field trial," *International Journal of Disaster Risk Reduction*, vol. 13, p. 307–312, 09 2015.

[12] M. Goodrich, B. Morse, D. Gerhardt, J. Cooper, M. Quigley, J. Adams, and C. Humphrey, "Supporting wilderness search and rescue using a camera-equipped mini uav: Research articles," *Journal of Field Robotics*, vol. 25, pp. 89 – 110, 02 2008.

[13] L. Lin, M. Roscheck, M. Goodrich, and B. Morse, "Supporting wilderness search and rescue with integrated intelligence: Autonomy and information at the right time and the right place." vol. 3, 01 2010.

[14] A. Birk, B. Wiggerich, H. Buelow, M. Pfingsthorn, and S. Schwertfeger, "Safety, security, and rescue missions with an unmanned aerial vehicle (uav)," *Journal of Intelligent & Robotic Systems*, vol. 64, pp. 57–76, 10 2011.

[15] P. Chrobocinski, E. Makri, N. Zotos, C. Stergiopoulos, and G. Bogdos, "Darius project: Deployable sar integrated chain with unmanned systems," in *2012 International Conference on Telecommunications and Multimedia (TEMU)*, July 2012, pp. 220–226.

[16] D. Serrano, G. De Cubber, G. Leventakis, P. Chrobocinski, D. Moore, and S. Govindaraj, "Icarus and darius approaches towards interoperability," 01 2015.

[17] M. T. Agcayazi, E. Cawi, A. Jurgenson, P. Ghassemi, and G. Cook, "Resquad: Toward a semi-autonomous wilderness search and rescue unmanned aerial system," in *2016 International Conference on Unmanned Aircraft Systems (ICUAS)*, 2016, pp. 898–904.

[18] J. P. Queralta, J. Taipalmaa, B. C. Pullinen, V. K. Sarker, T. N. Gia, H. Tenhunen, M. Gabbouj, J. Raitoharju, and T. Westerlund, "Collaborative multi-robot systems for search and rescue: Coordination and perception," *ArXiv*, vol. abs/2008.12610, 2020.

[19] S. Waharte and N. Trigoni, "Supporting search and rescue operations with uavs," in *2010 International Conference on Emerging Security Technologies*, 2010, pp. 142–147.

[20] N. H. Motlagh, M. Bagaa, and T. Taleb, "Uav-based iot platform: A crowd surveillance use case," *Comm. Mag.*, vol. 55, no. 2, p. 128–134, Feb. 2017.

[21] Y. Liu, H.-N. Dai, Q. Wang, M. K. Shukla, and M. Imran, "Unmanned aerial vehicle for internet of everything: Opportunities and challenges," *Computer Communications*, vol. 155, pp. 66 – 83, 2020.

[22] M. Satyanarayanan, "The emergence of edge computing," *Computer*, vol. 50, no. 1, pp. 30–39, 2017.

[23] Z. Zhou, X. Chen, E. Li, L. Zeng, K. Luo, and J. Zhang, "Edge intelligence: Paving the last mile of artificial intelligence with edge computing," *Proceedings of the IEEE*, vol. 107, no. 8, pp. 1738–1762, Aug 2019.

[24] M. Gao, W. Cui, D. Gao, R. Shen, J. Li, and Y. Zhou, "Deep neural network task partitioning and offloading for mobile edge computing," in *2019 IEEE Global Communications Conference (GLOBECOM)*, Dec 2019, pp. 1–6.

[25] C. Yang, Y. Liu, X. Chen, W. Zhong, and S. Xie, "Efficient mobility-aware task offloading for vehicular edge computing networks," *IEEE Access*, vol. 7, pp. 26 652–26 664, 2019.

[26] H. Ghazzai, A. Kadri, M. Ben Ghorbel, and H. Menouar, "Optimal sequential and parallel uav scheduling for multi-event applications," in *2018 IEEE 87th Vehicular Technology Conference (VTC Spring)*, 2018, pp. 1–6.

[27] Y. Wang, Z. Y. Ru, K. Wang, and P. Q. Huang, "Joint deployment and task scheduling optimization for large-scale mobile users in multi-uav-enabled mobile edge computing," *IEEE Transactions on Cybernetics*, vol. 50, no. 9, pp. 3984–3997, 2020.

[28] M.-A. Vasile, F. Pop, R.-I. Tutueanu, V. Cristea, and J. Kołodziej, "Resource-aware hybrid scheduling algorithm in heterogeneous distributed computing," *Future Generation Computer Systems*, vol. 51, pp. 61–71, 2015, special Section: A Note on New Trends in Data-Aware Scheduling and Resource Provisioning in Modern HPC Systems.

[29] A. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *ArXiv*, vol. abs/1704.04861, 2017.

[30] Q. Yang, X. Luo, P. Li, T. Miyazaki, and X. Wang, "Computation offloading for fast cnn inference in edge computing," 09 2019, pp. 101–106.

[31] A. Hashimoto and N. Abaid, "An agent-based model of lost person dynamics for enabling wilderness search and rescue," in *ASME 2019 Dynamic Systems and Control Conference*, 2019.

[32] R. K. Williams, A. Gasparri, and G. Ulivi, "Decentralized matroid optimization for topology constraints in multi-robot allocation problems," *2017 IEEE International*, 2017.

[33] Y. Sung, A. K. Budhiraja, R. K. Williams, and P. Tokekar, "Distributed assignment with limited communication for multi-robot multi-target tracking," *Auton. Robots*, 2020.

[34] J. Liu, L. Zhou, P. Tokekar, and R. K. Williams, "Distributed resilient submodular action selection in adversarial environments," *IEEE Robotics and Automation Letters*, vol. 6, no. 3, pp. 5832–5839, Jul. 2021.

[35] B. G. Cangan, L. Heintzman, A. Hashimoto, N. Abaid, and R. K. Williams, "Anticipatory human-robot path planning for search and rescue," *ArXiv*, vol. abs/2009.03976, 2020.

[36] L Heintzman, A Hashimoto, N Abaid, R K Williams, "Anticipatory planning and dynamic lost person models for Human-Robot search and rescue," in *IEEE International Conference on Robotics and Automation*, 2021.

[37] L. Heintzman and R. K. Williams, "Multi-agent intermittent interaction planning via sequential greedy selections over position samples," *IEEE Robot. Autom. Lett.*, vol. 6, no. 2, pp. 534–541, Apr. 2021.

[38] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," vol. 9351, 10 2015, pp. 234–241.

[39] Z. Cao, G. Hidalgo, T. Simon, S.-E. Wei, and Y. Sheikh, "Openpose: Realtime multi-person 2d pose estimation using part affinity fields," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 1, pp. 172–186, 2021.

[40] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," in *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, ser. AAAI'17. AAAI Press, 2017, p. 4278–4284.

978-1-6654-2369-4/21/$31.00 ©2021 IEEE