Realizing the Promise of Artificial Intelligence for Unmanned Aircraft Systems through Behavior Bounded Assurance

Prakash Sarathy
Aerospace Systems
Northrop Grumman Corp.
Redondo Beach CA
sriprakash.sarathy@ngc.com

Sanjoy Baruah
Computer Science & Engineering
Washington University
St. Louis MO
baruah@wustl.edu

Stephen Cook
Aerospace Systems
Northrop Grumman Corp
Melbourne, FL
stephen.cook@ngc.com

Marilyn Wolf
Electrical & Computer Engg.
Georgia Institute of Technology
Atlanta,GA
wolf@ece.gatech.edu

***Abstract*—A key value proposition for incorporation of Artificial Intelligence (AI) and Machine Learning (ML) methods into aviation is that they offer means of understanding data in ways that allow hitherto unprecedented insights for decision making, whether by a human or a machine. When these techniques are applied to cyber-physical systems, such as unmanned aircraft systems (UAS), they can result in positive societal impacts (e.g., search and rescue). However, the advantages of such techniques must be balanced against appropriate safety and security requirements so that taken together the system can ensure an acceptable level of confidence and assurance in both civilian and military applications. To this end, there is a need for the capability to suitably characterize such techniques and assess how they can be integrated into a viable assurance framework that can maximize safety and security benefits while bounding the inherent risk of non-determinism arising from such these approaches.**

This paper focuses on assurance and behavior bounds for decision making systems from a) algorithmic functional performance; b) schedulability analysis and candidate scheduling paradigms; and c) processor architectures (including multi-core) to support minimized interference in general. We will place particular emphasis on machine learning approaches for control, navigation and guidance applications for unmanned systems. This paper will review available and emerging approaches (e.g., formal methods, modeling and simulation, real-time monitors/agents among others) to ensuring behavior assurance for unmanned systems engaged in missions of moderate-to-high complexity. The intent is to examine behavior assurance for advanced autonomous operations within a holistic life-cycle process

Keywords—UAS behavior assurance, certification, avionics, AI/ML, algorithms, bounded behavior, guidance navigation and control.

I. INTRODUCTION

Data centric computing paradigms and approaches form the foundation of what is loosely regarded as the second coming of AI in recent years. [1-3] The significant upsurge of private and government investment in the area of AI is reminiscent of a similar trend in internet and telecom a decade or so back. Commercial applications range from driverless cars[4], aerial drones for package delivery[5], to undersea exploration[6], as well as surveillance and monitoring systems.

AI/ML techniques seem to offer the potential of key innovations in decision making [7-8], whether by a human or a machine. For cyber-physical systems, such as unmanned aircraft systems (UAS), AI/ML approaches offer the promise of capabilities and performance well beyond current state-of-practice. The convergence of processor architectures [10] and communications networks (5G) [11] coupled with the concepts of cloud and edge intelligence seem to be aligning with the specific needs of AI based applications. For the safety conscious aerospace industry, the AI/ML technologies need to be complimented by a strong assurance framework that can un-equivocally establish credibility and confidence in UAS that utilize them.

However, the potential upside of AI-based UAS must be balanced against appropriate safety and security requirements, so that taken together the system can ensure an acceptable level of confidence and assurance in both civilian and military applications. As with any major technological advance the underlying regulatory and certification processes lag behind and much work needs to be done in order to find the right balance of regulatory constraint and design freedom to innovate and operate. In particular, civilian airspace authorities are already burdened with aging infrastructure and overloaded air traffic management capacity in dealing with manned commercial aircraft. The introduction of un-manned aircraft into the mix creates additional challenges from both a capacity and regulatory standpoint. In particular, given the absence of any substantial prior work in the area of assuring UAS with advanced decision-making algorithms, the challenge is to develop new approaches to ensuring their safety within an integrated airspace. To this end, there is a need to adequately characterize AI-based techniques and assess how they can be integrated into a viable assurance framework that can maximize safety [12-14] and security benefits while bounding the inherent risk of non-determinism arising from such these approaches.

Figure 1: Example of State-of-Art autonomous aircraft with advanced avionics (Northrop Grumman X-47D)

A. AI/ML applications in UAS Avionics

Given the relatively nascent state of UAS avionics development and applicable mission capabilities, the target areas for initial insertion of AI/ML technologies are not well-defined. Although there a few instances of advanced autonomous systems incorporating this type of technology (Figure 1), most of the currently flying UAS are capable of only very simple missions. Nevertheless, based on the current successes of AI/ML in other domains the following key drivers applicable to UAS can be identified as:

Perception and modeling

a) Ability to extract perception information from large quantity of data
b) Ability to discern patterns and develop model structures

Open-ended problem solving

a) Reasoning under uncertainty
b) Decision making when conditions are "new"
c) Optimization under very large number of variables with unknown relationships

Potential to learn/discover new mechanisms

a) Ability to adapt to new situations
b) Ability to generate new approaches

These AI/ML capabilities can be quite relevant to developing advanced mission capabilities and ensuring a higher degree of resilience built-in to these UAS. Some immediate applications areas of these capabilities within a UAS framework are discussed here.

Mission Planning: Given that most available UAS have very simple planning capability which are either static or quite rigid in dealing with dynamic operational state changes, the use of Bayesian logic or statistical or stochastic approaches could be very useful in providing a significant advancement of planning capability. It is likely that these techniques will be applied initially to the path planning and route generation (guidance) problem for UAS.[15]

Obstacle Avoidance: A fundamental concern for safety for all UAS is the ability to sense and avoid obstacles while in flight and on the ground. AI/ML techniques are very well suited for enhancing extant obstacle avoidance algorithms in providing an un-precedented level of perception and awareness through learning-enabled components. [16,17]

Navigation and Control: Conventional control methods do not adapt well to changing operating conditions, therefore the use of adaptive and situation aware AI/ML approaches to high level navigation and control will significantly enhance the robustness of the UAS. [18,19]

Self-organization and Control: Effective command and control across a large aggregation of unmanned assets is a difficult problem. This challenge becomes many orders of magnitude more challenging when swarms are involved. AI/ML methods that can provide control of swarms through emergent behavior can be critical for such applications. [20,21]

B. Novel Approaches for Vehicle Navigation, Guidance, and Control

Focusing on a limited subset of ML/AI avionics applications described earlier, consider the safety critical task of flight guidance, navigation and control at the mission management level. Any algorithm that provides guidance (i.e a route or path) and/or navigation plan ("turn right at the next waypoint") and/or flight control inputs to a UAS vehicle management system is subject to the "flight/safety critical" certification criterion from a functional and temporal assurance standpoint. Examining two specific examples of emerging approaches for guidance, navigation and control (GNC) that utilize AI/ML approaches serve to better highlight the challenges ahead.

Backward Reachability Sets and Learning Safety: Work at UC Berkeley by Claire Tomlin and others [22,23], has focused on using machine learning methods to characterize and learn the implicit rules of safe operation for unmanned rotorcraft through the evolution of flight control (vertical descent) algorithms in online (in real-time) mode. The approach computes backward reachability sets to identify feasible paths that meet current "safety" characterization and metrics. On one hand this class of approaches take on the "safety-verifiability" problem head-on by building-in safety into the algorithms. However, this approach implies that the underlying flight-control laws are in a state-of-flux and are unknown a` priori! Current approaches to certification cannot handle this type of uncertainty and dynamic behavior.

Information theoretic Model Predictive Control (MPC): Theodorou et. al. at Georgia Tech [24,25] have developed information theoretic methods using MPC algorithms capable nonlinear optimization which is used to incorporate multi-layer artificial neural networks (ANN) as dynamics models. The MPC algorithms are used to solve model-based reinforcement learning tasks. This combination of MPC and ANN can be used in UAS to provide GNC implementation wherein there is no

overt control law to characterize. Again this poses a significant challenge to any conventional approach to certification under the safety critical provisions for UAS.

C. Assurance Characteristics of Advanced Mission Management Algorithms

As has been implied in earlier sections, a potential obstacle to largescale utilization of advanced algorithms based on data-centric approaches, is the challenge in establishing assurance arguments for this class of algorithms [26,27]. Although current certification procedures are not equipped to handle non-deterministic approaches, bounded behavior methods [28] offer feasible alternatives to managing both state-space explosion and the inherent uncertainty associated with data-centric methods in general and AI/ML techniques in particular. Characterizing data-centric methods is often a key challenge in developing behavior bounds as an initial step of Bounded Behavior Assurance (BBA) methodology [CITEXXX]. Some classes of AI/ML methods face additional challenges based on the specific mechanizations utilized in their implementation. A few examples of these challenges broken out by algorithm types are identified below.

Statistical and Stochastic Data-driven Methods: A large class of AI/ML algorithms are focused on feature classification and pattern-matching and many use some form of Bayesian estimators, Markov chains or Kalman filters to perform these functions. These techniques present challenges in establishing useful bounds (performance guarantees, convergence, a' priori error estimation) for use in BBA.

Neural Network Computing: Both conventional (ANN) and deep (DNN) neural networks provide unprecedented capabilities for a wide range of optimization and decision-making applications. However, almost without exception the quality of the resultant product is predicated heavily on the availability and integrity of training data sets and their sensitivity to design parameters, as well as the dependency of the performance to the implementation topology, provide significant obstacles to establishing a BBA based assurance profile for such approaches.

D. Implications for Cyber-Physical Assurance

AI workloads have different characteristics than do traditional algorithmic workloads: computationally intensive, may require accelerators, non-negligible latency. Autonomous AI workloads may be both sporadic and safety-critical. Traditional real-time workload models emphasize periodic execution with execution times within small variances. Understanding the environment may require reacting to changes to the environment, introducing sporadic computation. Acceleration and heterogeneous computation move scheduling and allocation from the traditional real-time systems model of tasks on closely-coupled processors to a loosely-coupled heterogeneous multiprocessor as in hardware/software co-design.

II. Certification Challenges for AI/ML based UAS

Enabling increased levels of autonomy for UAS through AI/ML has vast potential to benefit both safety and efficiency. The 2016 AIAA Intelligent Systems Technical Committee Roadmap for Intelligent Systems in Aerospace [29] concluded that incorporating adaptive features, such as those made possible by AI/ML can,

> *". . . improve efficiency, enhance performance and safety, better manage system uncertainty, as well as learn and optimize both short-term and long-term system behaviors."*

However, incorporation of AI/ML into UAS and aviation as a whole faces significant airworthiness certification challenges. While some of these challenges are not necessarily unique to UAS, the absence of a human pilot onboard the aircraft to provide the situation awareness, detect-and-avoid, and contingency management necessary to ensure safety of airspace users and people on the ground means that the software must implement all pilot-in-command intended functions necessary to ensure safety. The primary standard used for evaluating software aspects of safety and airworthiness for civilian aircraft is RTCA DO-178C, "Software Considerations in Airborne Systems and Equipment Certification".[30] Military airworthiness authorities sometimes use different standards for assuring software, but many of the objectives and activities are the same. While existing standards have proven effective for assuring software in today's airborne systems, there is some question as to whether they are scalable for AI/ML. Given the ability of AI/ML to "learn", the repeatability of test results – a key cornerstone of certification – becomes a challenge. Lacher et al. [31] observed that "Traditional mechanisms for exhaustive testing will not work for a system that may make different decisions given the same input with all of them potentially correct." Furthermore, a 2019 Forum for Aeronautical Software (FAS) report [32] notes several challenges with applying DO-178C to AI/ML:

1) Writing requirements for the software in a manner that include safety and security considerations for the intended function(s) (e.g., decomposing contingency management into testable, specific intended functions)
2) Understanding the role of the data inputs to the AI/ML in performance of the intended function(s) (e.g., accommodating unanticipated inputs into the software algorithms)
3) Understanding what the software actually contributes to the intended function relative to the data from which it is "learning" (e.g., is the code testable under all aircraft forseeable conditions)
4) Showing that the AI/ML implementation will return solutions within accepted bounds with an acceptable level of confidence (e.g., showing that AI/ML solutions will not be unsafe).

A. *Technology challenges with current certification processes particularly relevant to UAS*

Increasingly in aviation, and particularly in the small UAS industry, important factors in keeping software-related costs manageable involves the use of software re-use, open source code, and commercial-off-the-shelf components (some of which may contain embedded software). The small UAS industry has even embraced some open-source real-time operating systems – traditionally these operating systems are pedigreed to the highest level of software assurance. The FAS report [32] concluded that supplemental guidance is needed to "*… help the applicant better assess the effort to use/integrate Open Source/COTS software life cycle data in the final product.*" Implementing AI/ML solutions with components that are not assured with traditional methods calls into question aircraft airworthiness and the safety level of UAS operating in non-segregated airspace.

B. *Policy and standards challenges*

In addition to these certification challenges, there are key policy hurdles to integrating AI/ML onto UAS. Our current airspace regulatory system was written at a time when it was assumed that a human pilot-in-command was onboard the aircraft. For example, the U.S. Code of Federal Regulations [33] states:

§ 91.3 Responsibility and authority of the pilot in command.

a) The pilot in command of an aircraft is directly responsible for, and is the final authority as to, the operation of that aircraft.

b) In an in-flight emergency requiring immediate action, the pilot in command may deviate from any rule of this part to the extent required to meet that emergency.

c) Each pilot in command who deviates from a rule under paragraph (b) of this section shall, upon the request of the Administrator, send a written report of that deviation to the Administrator.

For a UAS implementing AI/ML to achieve functions normally reserved for a human pilot-in-command, it raises the question as to how this regulation applies to the AI/ML. If the human pilot-in-command is not onboard the aircraft and unable to intervene in the real-time decision making of the AI/ML, is the human still the "final authority"? Does the AI/ML have the authority to "deviate from any rule of this part to the extent required"? Further considerations of pilot training, liability in the event of mishap, and interaction with air traffic control will also have to be adjudicated.

Another potential barrier to the implementation of AI/ML surrounds the question of risk perception from the public. A 2018 study on public perception of autonomy [34] found "…an overwhelming response of uncomfortable feelings toward autonomy in aircraft…" among those surveyed. Despite evidence that automation has improved safety in aviation over the last several decades, many people remain skeptical of the introduction of autonomy into aviation. Since policy-makers and regulators must answer to the public, promoting trust in autonomy is a key aspect to the introduction of AI/ML.

C. *Moving beyond conventional certification methods for UAS integrated into civillian airspace*

For the aerospace industry to move forward with the introduction of AI/ML into UAS, several enablers are recommended. First of all, the recent FAA transition to performance-based standards should be expanded and applied to UAS in general and in particular to AI/ML. The FAA stated in the Notice of Proposed Amendment to Part 23 [35], "Incorporating the use of consensus standards as a means of compliance with performance-based regulations would provide the FAA with the agility to more rapidly accept new technology as it develops, leverage industry experience and expectations to develop of new means of compliance documents, and encourage the use of harmonized means of compliance…." One such consensus standard is ASTM F3269, "Standard Practice for Methods to Safely Bound Flight Behavior of Unmanned Aircraft Systems Containing Complex Functions" which provides for the use of a pedigreed run-time assurance (RTA) architecture to bound the behavior of an untrusted complex function (i.e., one enabled by AI/ML). [35] Figure 2 shows the F3269 generic RTA architecture which bounds the behavior of the complex function using a pedigreed safety monitor and pedigreed recovery control functions. This is one example of an implementation of bounded behavior assurance for UAS containing AI/ML.

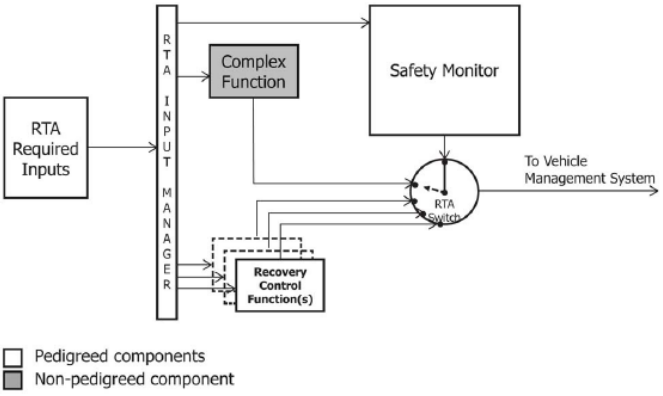

Figure 2. ASTM F3269 Generic Run-Time Assurance Architecture (from [36]).

Another key enabling activity for incorporation of AI/ML into UAS is to encourage policy-makers and regulators to build public trust in UAS containing AI/ML. The 2016 Defense Science Board Summer Study on Autonomy stated, "Establishing trustworthiness at design time and providing adequate capabilities so that inevitable variations in operational trustworthiness can be assessed and dealt with at run time is essential, not only for operators and commanders, but also for designers, testers, policymakers, lawmakers, and the American public." [36] Research into methods to assure AI/ML in the design are ongoing and should continue. A 2019 ASTM

Technical Report [37] recommends tailoring the airworthiness requirements and means of compliance for autonomy by taking into account the role of the automation, the complexity of the automation, and the net risk of incorporating automation into the aircraft design versus having the human pilot perform the function. Operational trustworthiness can be gained through a "crawl, walk, run" approach where UAS are first permitted in low-risk operational areas and then expanded outside of these safety corridors, moving into more integrated airspace as trust is built. It is important that the risk perception from all aviation stakeholders – manufacturers, pilots, air traffic controllers, regulators, etc. – be consulted and included as operations expand.

Finally, policies and regulations should take into account the benefits of UAS with AI/ML and not only focus on the risks. A 2018 National Academy of Sciences report [38] recommended, "UAS operations should be allowed if they decrease safety risks in society – even if they introduce new aviation safety risks – as long as they result in a net reduction in total safety risk." By using a bounded behavior framework approach to integrating AI/ML into UAS, it is feasible to realize the benefits of this new technology for aviation, provide a path to airworthiness certification, build trust with policy-makers, regulators, and the public, and increase overall safety to the public.

III. The Bounded Behavior Assurance (BBA) Methodology

The Bounded Behavior Assurance (BBA) approach [28] to assuring safety of autonomous systems is centered upon being able to specify *bounds* on the functional and timing run-time aspects of system prior to run-time, such that safety is assured so long as the state remains within these bounds. During run-time, decisions are taken to optimize mission-effectiveness while simultaneously *monitoring* system state [39] to determine if behavior is approaching the boundaries of safety. If this happens, mission-effectiveness ceases to become the primary objective; instead *corrective action* is taken to ensure that system state remains within the acceptable bounds. Hence the BBA approach does not seek the eliminate the inherent non-determinism and unpredictability of run-time behavior in AI/ML-based autonomous systems; instead, it uses the mechanism of run-time monitoring (and corrective action, if needed) in order to bound the degree of such unpredictability.

It is becoming increasingly clear that safety assurance in autonomous systems cannot realistically be looked upon as a single-step process that is completed after system development and prior to deployment; instead, a lifelong approach is required that must be initiated at the very beginning of the design process and continue through the lifetime of the system. The BBA approach to safety assurance requires that acceptable safe system states be identified and formally specified[40]. Doing so becomes a serious challenge as the functional requirements placed upon autonomous systems repeatedly evolve during the life-time of the system: novel use-case scenarios that were unanticipated at design time and initial deployment are frequently envisioned and proposed, and added on to the mission-capability requirements for the autonomous systems. The open-ended nature of many planned operational scenarios for autonomous systems inevitably means that these systems will confront completely unanticipated scenarios — "**unknown unknowns**": what kinds of assurance requirements could be placed upon system behavior in such unanticipated scenarios?

Although the BBA approach to safety assurance offers a promising alternative to conventional testing-based approaches (which appear to not generalize easily to complex autonomous systems that incorporate learning), it does require a rigorous cross-layer approach to system specification and analysis that incorporates integrated consideration of functional and timing behavioral issues at the levels of hardware, the real-time operating system (RTOS), scheduler and associated resource-allocation mechanisms, and the application layer. These and related open issues must be thoroughly investigated before the BBA approach can be fully implemented as a means of assuring safety for complex autonomous systems.

A. Functional Assurance

AI/ML based algorithms have numerous different roles in autonomous systems [41]. In *perception,* they are responsible for reading in input from physical sensors such as cameras and determining what that input represents. ("Smart" cameras, such as the ones used in driver-assist features in cars for identifying pedestrians, are an example of this use of AI/ML in an autonomous system.) For such use in perception, it is often the case that exact formal specifications are simply not available: there is no formal machine-checkable specification for the field of pixels that distinguish a pedestrian from, e.g., a color photograph of a human being. Hence it appears that assurance cases for AI/ML-based perception systems must be based on extensive testing, and assurance arguments explaining why the training data used to train the system should be considered adequate. The development of a framework for such assurance cases is one of the major open issues that need to be addressed. In addition to perception, AI/ML based algorithms may play an important role in *reflection*: incorporating input from perceptors, other sensors, and prior knowledge in order to develop an internal understanding of the environment within the system is operating. Here again, there are major challenges that must be addressed in order to be able to apply the BBA approach for assuring system safety. How do bounds on the uncertainties in perception from multiple different sources compose with each other, along with bounds on the uncertainties in prior knowledge, to provide the safety envelope within which run-time behavior is allowed to exist?

AI/ML based systems are also used for *decision making* by autonomous systems. Additional challenges must be overcome in order to render BBA applicable to such use of AI/ML; in particular, formal methods are needed that allow for the specification of behavior that incorporates bounds on acceptable system state, and techniques must be developed that allow for mapping bounded-behavior models of the external environment (conditions; threats; etc.) onto these bounded-

behavior models of the output produced by the AI/ML based decision-making component.

B. Temporal Assurance

Requirement specifications for unmanned aircraft systems have a temporal component in additional to a functional one: specified functional outputs are required to be computed within specified latencies. Establishing beforehand that such temporal specifications will always be met during all executions of the system is already challenging for current (non-autonomous) systems due to the high degree of timing unpredictability inherent in modern commercial off-the-shelf (COTS) components that comprise the computing platforms in modern aircraft; these challenges are further exacerbated when the required functions are computed using AI/ML based algorithms for which the functional outputs are also unpredictable. Novel scheduling and schedulability-analysis approaches are being explored for enabling the application of the BBA approach for assuring timing correctness in autonomous systems using AI/ML based algorithms. Some such approaches break down a functional computation for which an end-to-end delay bound must be assured as a multi-stage computation in which a sequence of functional blocks must complete execution within a specified duration. During run-time both the duration of execution of a stage, and and some estimate of the quality of the functional output produced by that stage, are monitored at the end of each stage. These monitored values guide the choice of implementation for the following functional blocks: if the duration that has elapsed during some particular stage is larger than expected then a simpler implementation choice that has a smaller expected execution duration must be selected for the following stages in order to ensure timing safety. Conversely if some stage completes execution sooner than expected then a more sophisticated computation-intensive AI/ML based implementation can be selected for the following stage, which allows for improved performance of the mission of the autonomous system without compromising safety. A priori characterization of the expected execution duration of each implementation, and of the value returned by it, may be deterministic or stochastic; if the latter, then probabilistic schedulability analysis techniques are used.

C. Processor Architecture Considerations – Interference/ Isolation

Embedded computing and cyber-physical systems must satisfy not only functional requirements but also timing, power, and thermal behavior. Functional requirements describe input/output behavior. Timing behavior is often in the form of latency from input to output. Power consumption for a given computation is limited both by the capacity of the associated power supply. Exceeding certain temperatures governed by device physics will result in physical damage to the processor; thermal behavior can serve as an indirect limit on power consumption. All of these non-functional requirements require complex analysis for modern processors [42]. Processor microarchitectures, caches, and memory systems all contribute complexity to the modeling of timing/power/thermal behavior. In most cases for modern processors, we can only estimate these properties within certain bounds. The characteristics of software also add to the complexity of non-functional analysis. Complex execution paths, data dependencies, data and code placement in the cache, and other factors all make estimation of software characteristics difficult.

Many system-on-chip applications exhibit non-uniform workloads that can cause variations in computational load, power consumption, and thermal load. These variations can occur over both time and space. Temporal variations may result in metaphorical hot spots of heavy computation on processing elements or communication load for the network-on-chip. Spatial variations may result in literal hotspots with elevated temperatures on some parts of the chip. Embedded computer vision is an example application with non-uniform workloads. Xu *et al.* [43] measured the workload created by a real-time gesture recognition system and developed a design methodology for application-specific, non-regular networks-on-chips tailored to a given traffic pattern.

Digital neural networks (DNNs) can be optimized to reduce computational, memory bandwidth, and power requirements. Networks can be optimized using several methods. Reduced-precision arithmetic can result in smaller bit widths for values and smaller arithmetic units. Zhuang *et al.* [44] progressively decreased bit width during training; they also used a full-precision model in parallel to guide training of the limited-precision model. Another approach is creating sparse coefficient networks that reduce memory bandwidth thanks to zero coefficients that do not need to be fetched. Guo *et al.* [45] compressed DNNs using structured sparsity learning. Li *et al.* [46] applied structured sparsity learning as a co-design methodology.

Several CPU attacks have recently been identified that compromise supposedly protected data during execution. The Meltdown vulnerability [47], discovered in 2017 and announced in 2018, exploits a race condition and a cache side-channel attack to avoid checks that are meant to enforce isolation between processes. Spectre [48], also discovered in 2017 and announced in 2018, exploits effects of speculative execution due to branch misprediction. The Microarchitectural Data Sampling (MDS) attack [49], announced in 2019, refers to a set of techniques related to side channel effects from speculative execution. We can identify several threat models for the disclosure of data from the computing platform:

- Data may be directly observed using circuits attached to the computing platform. An attacker with physical access to the system may be able to attach monitoring devices.
- Side channel attacks may observe data using electromagnetic radiation, power measurements, or other physical mechanisms.

Several threat models for software stack target both functional [50] and non-functional behavior:

- Trojan horses or tampering with software installation can introduce malicious code that can both observe and modify behavior.
- Timing attacks change the timing behavior of the system without necessarily modifying values. A missed deadline can have serious effects.
- Replay attacks [51, 52] send previously recorded data to other parts of the system to hide their changes to the activity of signals. Replay attacks were used in the Stuxnet attack to hide software attacks that damaged equipment.

IV. Conclusions and Recommendations

Incorporation of AI/ML into UAS and aviation at large offers potential benefits to performance, cost, and safety. However, there are immediate challenges that accompany the use of AI/ML methods and hurdles that may prevent the realization of the promise of these technologies for UAS applications. We examined promising AI/ML approaches that have near term relevance for UAS avionics in general and focused on guidance, navigation and control applications. We summarize some of the key considerations for advanced approaches to mission management algorithms for UAS. We have highlighted some of the fundamental assurance challenges inherent in these approaches for safety critical applications. We then illustrated specific challenges in current certification approaches and associated regulation and policy governing UAS in civilian airspace while noting the changing direction of regulatory authorities in response to these challenges.

We also presented an overview of an assurance approach (BBA) that can be used to develop assurance framework for UAS using advanced technologies such as AI/ML algorithms Using BBA or similar approaches still requires significant advances in maturing and characterizing these algorithms vis-à-vis safety and security. We proceed to identify some of these obstacles from both functional and temporal assurance perspectives.

We conclude this review by examining the underlying security and safety implications for processor hardware architectures that are likely to be used to drive the computational loads associated with AI/ML methods. We find that a BBA approach toward incorporating AI/ML into UAS has the potential to accelerate the integration of these technologies into the aircraft. Furthermore, we find that a BBA approach has benefits with regard to moving UAS from lower risk segregated operations to more operationally relevant parts of the airspace. We recommend that developers, regulators, policymakers, and researchers evaluate the BBA approaches and use them to the maximum extent to realize the promise of artificial intelligence and machine learning in UAS.

References

[1] Jim Sinur, “Is AI Data Driven, Algorithm Driven, or Process Driven?” in COGNITIVE WORLD, https://www.forbes.com/sites/cognitiveworld/

[2] Kai-Fu Lee, “The Four Waves of A.I.”, in FORTUNE, October 22, 2018, https://fortune.com/2018/10/22/artificial-intelligence-ai-deep-learning-kai-fu-lee/

[3] Christoffer Tønnessen & Sigve Farstad, “The Second Coming of Artificial Intelligence”, March, 2017. http://blog.telenor.io/2017/03/03/The-Second-Coming-Of-Artificial-Intelligence.html

[4] Lipson, H. and Kurman, M., Driverless: Intelligent Cars and the Road Ahead, MIT Press, 2016, ISBN 9780262035224

[5] Sarah E. Kreps, Drones: What Everyone Needs to Know. Oxford University Press, 2016. ISBN: 0190235373, 97801902353

[6] Daniel R. Faust, Underwater Robots. The Rosen Publishing Group, Inc, 2016. ISBN 1499421893, 9781499421897

[7] Editors:Gloria Phillips-Wren and Nikhil Ichalkaranje, “Intelligent Decision Making: An AI-Based Approach”, Volume 97 of Studies in Computational Intelligence, Springer Science & Business Media, 2008

[8] Casey C.Bennettab and Kris Hauserb., “Artificial intelligence framework for simulating clinical decision-making: A Markov decision process approach” in Artificial Intelligence in Medicine, Volume 57, Issue 1, January 2013, Pages 9-19

[9] Robert R. Trippi and Efraim Turban (Eds.)., “Neural Networks in Finance and Investing: Using Artificial Intelligence to Improve Real World Performance”. McGraw-Hill, Inc., New York, NY, USA. 1992.

[10] Hwang, K., Ghosh, J., and Chowkwanyun, R., Computer architectures for artificial intelligence processing. United States: N. p., 1987. Web. doi:10.1109/MC.1987.1663352.

[11] Chavez-Santiago, R, Szydelko, M, Kliks, A, Foukalas, F, Haddad, Y, Nolan, KE, Kelly, MY, Masonta, MT and Balasingham, I., “5G: the convergence of wireless communications”. Wireless Personal Communications, vol. 83(3), 2015. pp. 1617-1642

[12] Yampolskiy, R. (Ed.). (2018). Artificial Intelligence Safety and Security. New York: Chapman and Hall/CRC, https://doi.org/10.1201/9781351251389

[13] James Babcock, Janos Kramar, Roman V. Yampolskiy, “Guidelines for Artificial Intelligence Containment”. 2017. arXiv:1707.08476 [cs.AI]

[14] SA Seshia, D Sadigh, SS Sastry , “Towards verified artificial intelligence.”- arXiv preprint arXiv:1606.08514, 2016 - arxiv.org

[15] Michael Brady, “Artificial intelligence and robotics”, Artificial Intelligence Volume 26, Issue 1,1985, Pages 79-121, ISSN 0004-3702, https://doi.org/10.1016/0004-3702(85)90013-X.

[16] Kistan, T.; Gardi, A.; Sabatini, R. “Machine Learning and Cognitive Ergonomics in Air Traffic Management: Recent Developments and Considerations for Certification.” *Aerospace* **2018**, *5*, 103.

[17] E. Blasch, "Autonomy in Use for Information Fusion Systems", *NAECON 2018 - IEEE National Aerospace and Electronics Conference*, Dayton, OH, 2018, pp. 1-8. doi: 10.1109/ NAECON. 2018.8556777

[18] H. Thi Tu Uyen, P. Duc Tuan, L. Viet Anh and P. Xuan Minh, "Adaptive Neural Networks Sliding Mode Backstepping Control for 3DOF Surface Ship with Uncertain Model," *2018 International Conference on System Science and Engineering (ICSSE)*, New Taipei, 2018, pp. 1-6. doi: 10.1109/ICSSE.2018.8520227

[19] Shin, J., Kim, S. & Tsourdos, A, “Neural-networks-based Adaptive Control for an Uncertain Nonlinear System with Asymptotic Stability.” in Int. J. Control Autom. Syst. (2018) 16: 1989. https://doi.org/10.1007/s12555-017-0641-x

[20] A. Vega, A. Buyuktosunoglu and P. Bose, "Towards “Smarter” Vehicles Through Cloud-Backed Swarm Cognition" in *2018 IEEE Intelligent Vehicles Symposium (IV)*, Changshu, 2018, pp. 1079-1086. doi: 10.1109/IVS.2018.8500627

[21] Patel Y.S., Misra R., Mishra M.K., Mishra B.S.P. “Intelligent Computational Techniques for the Better World 2020: Concepts, Methodologies, Tools, and Applications.” In: Mishra M., Mishra B., Patel Y., Misra R. (eds) Smart Techniques for a Smarter Planet. Studies in Fuzziness and Soft Computing, vol 374. Springer, Cham 2019

[22] Humberto Gonzalez, Ram Vasudevan, Maryam Kamgarpour, S. Shankar Sastry, Ruzena Bajcsy, and Claire J. Tomlin. “A descent algorithm for the optimal control of constrained nonlinear switched dynamical systems.” In Proceedings of the 13th ACM international conference on Hybrid systems: computation and control (HSCC '10). ACM, New York, NY, USA, 51-60. 2010

[23] J. F. Fisac, A. K. Akametalu, M. N. Zeilinger, S. Kaynama, J. Gillula and C. J. Tomlin, "A General Safety Framework for Learning-Based Control in Uncertain Robotic Systems," in *IEEE Transactions on Automatic Control*, vol. 64, no. 7, pp. 2737-2752, July 2019.

[24] Williams, Grady & Wagener, Nolan & Goldfain, Brian & Drews, Paul & Rehg, James & Boots, Byron & A. Theodorou, Evangelos. (2017). "Information theoretic MPC for model-based reinforcement learning." . *2017 IEEE International Conference on Robotics and Automation (ICRA)*. 1714-1721. 10.1109/ 7989202.

[25] G. Williams, B. Goldfain, P. Drews, J. M. Rehg and E. A. Theodorou, "Best Response Model Predictive Control for Agile Interactions Between Autonomous Ground Vehicles," *2018 IEEE International Conference on Robotics and Automation (ICRA)*, Brisbane, QLD, 2018, pp. 2403-2410. doi: 10.1109/ICRA.2018.8462831

[26] Dreossi T. et al. (2019). VerifAI: A Toolkit for the Formal Design and Analysis of Artificial Intelligence-Based Systems. In: Dillig I., Tasiran S. (eds) Computer Aided Verification. CAV 2019. Lecture Notes in Computer Science, vol 11561. Springer, Cham

[27] Weiming Xiang, Taylor T. Johnson, "Reachability Analysis and Safety Verification for Neural Network Control Systems." 2018. arXiv:1805.09944 [cs.SY]

[28] Lee, J., Prajogi, A., Rafalovsky, E., and Sarathy, P. "Assuring behavior of autonomous UxV systems." In S5: The Air Force Research Laboratory (AFRL) Safe and Secure Systems and Software Symposium, July 2016.

[29] "Roadmap for Intelligent Systems" in Aerospace. American Institute of Aeronautics and Astronautics Intelligent Systems Technical Committee. First Edition. June 6, 2016.

[30] RTCA DO-178C, "Software Considerations in Airborne Systems and Equipment Certification," Washington DC, 13 December 2011.

[31] Lacher, A., Grabowski, R., and Cook, S. "Autonomy, Trust, and Transportation", in 2014 AAAI Spring Symposium Series, 2014.

[32] Forum for Aeronautical Software Unmanned Air System Ad-Hoc Group, "Internal Report on the Applicability of the DO-178C/ED-12C Related Set of Software Documents for the Development of UAS Software," RTCA Paper No: 054-19/PMC-1859, 25 February 2019.

[33] United States Code of Federal Regulations, 14 CFR 91.3 "Responsibility and authority of the pilot in command." As of 09 July 2019.

[34] Wollert, Matthew, "Public Perception of Autonomous Aircraft," Arizona State University, May 2018.

[35] "Revision of Airworthiness Standards for Normal, Utility, Acrobatic, and Commuter Category Airplanes," Federal Aviation Administration Notice of Proposed Rulemaking, Docket No. FAA-2015-1621, Notice No. 16-01. March 14, 2016.

[36] ASTM F3269, "Standard Practice for Methods to Safely Bound Flight Behavior of Unmanned Aircraft Systems Containing Complex Functions," West Conshohocken PA, ASTM International, 2017.

[37] ASTM International, Autonomy Design and Operations in Aviation: Terminology and Requirements Framework, West Conshohocken PA, ASTM International, 2019.

[38] National Academy of Sciences, Engineering, and Medicine, Assessing the Risks of Integration Unmanned Aircraft Systems into the National Airspace System, Washington DC, The National Academies Press, 2018.

[39] Kane, A., Fuhrman, T., , and Koopman, G., Monitor based oracles for cyber-physical system testing: Practical experience report, in Dependable Systems and Networks (DSN), 2014 44th Annual IEEE/IFIP International Conference on, June 2014, pp. 148–155.

[40] High Dimensional Reachability Analysis: Addressing the Curse of Dimensionality in Formal Verification, Chen, Mo. University of California, Berkeley, ProQuest Dissertations Publishing, 2017. 10607827.

[41] Sifakis, J. Autonomous Systems – An Architectural Characterization. Models, Languages, and Tools for Concurrent and Distributed Programming 2019: 388-410

[42] Marilyn Wolf, High Performance Embedded Computing: Applications in Cyber-Physical Systems and Mobile Computing, second edition, Elsevier, 2014.

[43] Jiang Xu, Wayne Wolf, Joerg Henkel, and Srimat Chakradhar, "A design methodology for application-specific networks-on-chip," ACM Transactions on Embedded Computing Systems, 5(2), May 2006, pp. 263-280.

[44] B. Li, W. Wen, J. Mao, S. Li, Y. Chen, H. H. Li, "Running sparse and low-precision neural network: where algorithm meets hardware," in 2018 23rd Asia and South Pacific Design Automation Conference (ASP-DAC), 2018, pp. 534-539.

[45] Yiwen Guo, Anbang Yao, and Yurong Chen, "Dynamic network surgery for efficient DNNs," in 30th Conference on Neural Information Processing Systems (NIPS 2016), 2016.

[46] B. Zhuang, C. Shen, M. Tan, L. Liu, and I. Reid, "Toward effective low-bitwidth convolutional neural networks," 30th Conference on Neural Information Processing Systems (NIPS 2016), 2016, pp. 7820-7928.

[47] Intel, "Side Channel Vulnerability Microarchitectural Data Sampling," accessed May 17, 2019, https://www.intel.com/content/www/us/en/architecture-and-technology/mds.html?wapkw=speculative+attack

[48] Moritz Lipp, Michael Schwarz, Daniel Gruss, Thomas Prescher, Werner Haas, , ers Fogh, Jann Horn, Stefan Mangard, Paul Kocher, Daniel Genkin, Yuval Yarom, and Mike Hamburg, "Meltdown: Reading Kernel Memory from User Space," in 27th USENIX Security Symposium (USENIX Security 18), 2018.

[49] Paul Kocher, Jann Horn, , Anders Fogh, Daniel Genkin, Daniel Gruss, Werner Haas, Mike Hamburg, Moritz Lipp, Stefan Mangard, Thomas Prescher, Michael Schwarz ,and Yuval Yarom, "Spectre Attacks: Exploiting Speculative Execution," in 40th IEEE Symposium on Security and Privacy (S&P'19), 2019.

[50] Stephen Checkoway, Damon McCoy, Danny Anderson, Brian Kantor, Hovav Shacham, Stefan Savage, Karl Koscher, Alexei Czeskis, Franziska Roesner, and Tadayoshi Kohno, "Comprehensive Experimental Analyses of Automotive Attack Surfaces," in Proceedings of the USENIX Security Symposium, San Francisco, CA, August 2011.

[51] Nicholas Falliere, "Stuxnet introduces the first known rootkit for industrial control systems," Symantec Official Blog, August 6, 2010, http://www.symantec.com/connect/blogs/stuxnet-introduces-first-known-rootkit-scada-devices.

[52] Nicholas Falliere, Liam O Murchu, and Eric Chien, W32.Stuxnet Dossier, version 1.4 February 2011, available at www.symantec.com.