Technology Innovation Institute

BUILDING SAFER COLLECTIVES OF AUTONOMOUS ROBOTS WITH SECURE RUNTIME ASSURANCE

Runtime assurance (RTA) architectures apply redundant interconnected controllers to secure critical assets. The aerospace industry relies on RTA to reduce costs and increase the reliability of aircraft, satellites, and spacecraft. Over time, as costs have fallen and best practices have emerged, engineers have increasingly explored RTA for autonomous vehicles and other safety-critical systems. Fully secure runtime assurance (SRTA) could help scale traditional runtime systems to accommodate the growing ranks of autonomous systems managing a fleet of robots (e.g., swarms of UAVs, UGVs) —while providing security, resilience, and safety which will govern the deployment of autonomous transportation.

Traditionally, runtime assurance architectures have protected individual things and relied on relatively simple control systems. The conventional approach may, however, fail to protect multiple devices working together—whether in swarms of drones, fleets of vehicles, or autonomous embedded systems in factories, buildings, and smart cities. Secure RTA includes zero-trust architecture and supports machine learning. Zero-trust architectures integrate best practices for security and resilience into multiple levels of the hardware and software stacks. Adding support for artificial intelligence (AI) and its machine learning (ML) subset across these systems further improves the swarms' performance, safety, resilience, and security.

For example, after a technical glitch in July 2023, a light show of 500 drones in Melbourne, Australia suffered a dramatic black-out ending when as many as 440 of them tried to make emergency landings in the city's Yarra River [1]. Almost 350 were lost in the water. This was the largest such incident, but by no means the first. In 2022, for example, fifty drones were similarly lost in Perth, on Australia's west coast.

Fortunately, these incidents did no harm to people or property (other than the drones, of course). Still, these incidents illustrate the need for better runtime safety. Common sense tells the human operator not to land an electric drone in the water. Drone control systems require specific design instructions. Redundant control systems reinforced by machine learning and AI could have reduced this risk through improved anomaly detection, more adaptable communication and control mechanisms, and more complex safety protocols.

The Secure Systems Research Center (SSRC) of the Technology Innovation Institute (TII, Abu Dhabi, UAE, tii.ae) has spearheaded research into applying SRTA to drone fleets. Unmanned Aerial Vehicles (UAVs, commonly called drones) are an ideal starting point for designing and testing the critical components of SRTA, thanks to their relatively low cost, their mobility, and even their susceptibility to several emerging security and control risks. In the long run, the architectures, software, and hardware created for drone programs could also enhance the safety of a much wider variety of autonomous systems, swarms, and autonomous and collaborative AI control systems.

Note: *While some publications refer to autonomous "swarms," the term "fleets" is increasingly used in operational contexts. This paper follows the latter practice.*

EVOLUTION OF RTA

SRTA builds on RTA, itself developed to improve safety in aerospace, automotive, and nuclear power industries—wherever the cost of failure is exceptionally high. Lee et al. introduced the term "runtime assurance" in 1999, as part of their process for characterizing correctness of execution at run-time. [2]

In 2001, Sha elaborated on a framework that emphasized simplicity for controlling complexity in large software systems. [3] Subsequent research explored RTA's potential for complex cyber-physical systems. [4] Fuller et al. have also examined the rise of RTA architecture more broadly. [5]

Technical advances—in computing, sensors, and control systems—have expanded RTA's capabilities, but at the cost of additional complexity. The very latest RTA implementations take advantage of even more advanced sensor technologies, which allow for more comprehensive system monitoring and enhanced safety mechanisms. But changes in software development, especially applications based on more complex, non-deterministic algorithms, pose new challenges and require RTA systems to grow ever more adaptable and sophisticated.

R&D teams have expended considerable effort to improve RTA by embedding AI and ML.[6] The combination of advanced AI models, robust sensor networks, and adaptive control mechanisms promises to enhance the drones' ability to predict and adapt in diverse operational environments.

RTA is also being customized to meet drones› unique operational needs and challenges—focusing on autonomy, functional safety, and cybersecurity. As drones gain more autonomy in more applications, RTA has become crucial to ensuring their security and functional integrity. This will be critical to a safe rollout of drones that operate Beyond Visual Line of Sight (BVLOS).

For example, GE's Trusted Autonomy Project is developing sophisticated airborne collision avoidance algorithms for small, unmanned aircraft, leveraging Deep Neural Networks (DNNs) for essential decision-making processes. [7] This project underscores the potential of DNNs to facilitate safe operations out of visual range. Researchers are exploring BVLOS technology for applications in self-driving vehicles, defense, and space. GE's work also includes a formal verification framework that codifies metrics for the reliability of safety-critical systems that may be controlled by less-trusted applications.

Edge Case Research has also developed an RTA framework for autonomous vehicles and robotics, based on the UL4600 safety standard. [8] The US Air Force Research Lab is also developing an RTA framework for securing AI-based autonomous systems, focusing on collision avoidance. [9] BAE Systems' Trusted Autonomy initiative blends human and machine decision-making, starting with personnel carriers for which human crews can be optional. [10]

Developers also report progress towards advanced RTA components that decrease costs. For example, Auterion's Skynode for drones integrates a sophisticated flight controller running lower-cost, open-source software that supports autonomy and AI integration in regulated environments. [11] The SSRC is also developing similar modular hardware and software components on top of open RISC-V chip designs that additionally support zero-trust architectures (more on these below). [12]

CHALLENGES AND LIMITATIONS OF RTA

Despite this recent progress, RTA systems still face many challenges and limitations, particularly when designers wish to incorporate them in small, cost-effective drones and drone swarms:

RTA Implementation

- **Complex Integration with AI and ML:** AI and ML systems require RTA to accurately monitor, analyze, and respond to AI-driven decisions in real-time. This becomes increasingly daunting with non-deterministic AI models, of which there are many.

- **Variable Operational Environments:** Drones operate in diverse environments, from urban landscapes to remote, unstructured settings. RTA systems must be robust and adaptable to handle this wide range of conditions, which in turn requires advanced sensing and data processing capabilities, and the ability to adapt quickly to new or unforeseen circumstances.

Limitations of Existing RTA Systems

- **Computational Constraints:** RTA demands real-time processing to ensure safety. More capable processors will be required to handle the vast data volumes generated by advanced sensors. This is particularly challenging given drones' constraints on size, weight, and power.

- **Continuous Learning and Adaptation:** More flexible RTA systems must adapt and learn from complex operational environments to improve over time. This adds additional complexity, especially given demands for extensive validation and verification.

Challenges in Fleet Operations

- **Communication Among Drones:** Maintaining consistent and reliable communication among drones is crucial for real-time data-sharing and collective decision-making. In a complex or hostile environment, communication links can be unstable or degraded by interference, posing significant risks to operational integrity.

- **Scalability of RTA Systems:** As the size of the fleet increases, the complexity of monitoring each drone and assuring its safety grows exponentially. Scalability will require some combination of faster chips and more efficient algorithms.

- **Autonomous Decision-Making within Fleets:** Collective decision-making can improve drone fleets' trustworthiness, safety, security, and resilience. More sophisticated algorithms are required to balance individual autonomy with collective behavior.

Regulatory and Ethical Constraints

- **Regulatory Compliance:** More advanced RTA systems, particularly those that use AI, will need to adhere to an emerging web of regulations and ethical considerations, especially regarding privacy, data security, and liability.

- **Accountability in Swarm Operations:** Drone fleets raise new regulatory and ethical concerns, particularly regarding operators' accountability and responsibility for potential unintended consequences in complex interactions. RTA systems must address these concerns while maintaining operational efficiency and safety.

TRUSTED AUTONOMY

Trusted autonomy, characterized by advanced AI and collaborative teaming, lets systems execute on their own, reliably, and ethically. It also builds a foundation for mutual trust between autonomous entities and human operators, where machines develop reliable trust in human directives and vice versa. Trusted autonomy will also require novel formal verification techniques to validate the safety of complex algorithms, especially those driven by deep learning.

Progress requires developing a framework (with associated hardware and software stacks) that combines the advantages of trusted systems with AI, ML, and autonomous systems. This framing of secure runtime environments suggests that trust autonomy could help contextualize the integration of these previously disparate disciplines.

The auto industry has developed consensus definitions of "autonomous systems" that specify multiple levels of reliability and control, and other domains have adopted the approach. Trusted autonomy characterizes some of the ways SRTA (remember, this is secure runtime assurance) could enhance the safety and security of individuals and swarms. "Trusted autonomy" suggests a sophisticated leap beyond traditional autonomy.

Traditional autonomy emphasizes independent decision-making and goal achievement by autonomous systems. In contrast, trusted autonomy advances this concept by integrating trust and reliability into decision-making to enhance safety, efficacy, and ethical responsibility in complex operational environments.

Implementing trusted autonomy in autonomous systems encompasses intricate challenges around algorithmic complexity, establishing trust between humans and machines, assuring safety and reliability, and evolving these systems in response to changing conditions. Cybersecurity is critical, too, since these systems are vulnerable to external threats. Addressing these challenges is crucial for successfully deploying and operating trusted autonomous systems.

Assuring safety and reliability becomes more complex as the number of units increases. Again, this requires robust algorithms capable of real-time adaptation and fault tolerance. Trusted autonomy, especially in the context of swarm robotics and drone fleets, requires advanced strategies in trust establishment, collaborative autonomy, and dynamic system management. Addressing these challenges is essential for securely and reliably harnessing these systems' full potential.

Trusted autonomy and SRTA need to take into account how and how much innovations in AI and machine will strengthen RTA. Developers will need to consider extending trust to AI using zero-trust architecture principles. The zero-trust approach has traditionally focused on malicious threats, rather than pitfalls of collaborations. AI and autonomous systems also require new components in establishing trust, particularly given AI's tendency to hallucinate and the increasing dependence on sensors available in their surroundings, sensors that may not accurately measure their environments.

Scaling trust across autonomous swarms will require new architectures, approaches, and algorithms. Zero trust architecture could help combat wireless network attacks. [13] Go up one organizational level, and similar principles must be extended to autonomous systems in general. [14] Additionally, applying concepts of autonomic computing to autonomic swarms could scale trust to a system-of-systems level through multiple tiers of control and redundancy. [15]

THE VALUE OF AI AND ML IN RTA

Integrating AI and ML with RTA markedly expands the size of the datasets RTA systems can handle and expands their ability to identify patterns and anomalies. This analytical foresight is crucial in preemptively mitigating risks associated with system failures, unauthorized intrusions, and cybersecurity threats, opening the door to immediate, intelligent responses to ever-shifting situations.

Innovations in AI and ML have been game-changers in drone technology. "AI" covers any machine performing tasks that typically require human intelligence. "ML," a subset of AI, focuses on algorithms that learn and make data-based decisions from past experience. Together, AI and ML have significantly boosted drone autonomy, decision-making, and operational efficiency.

In drones, AI and ML are essential to navigation and autonomous flight. They give drones the power to independently traverse complex environments, driving intricate path planning, sophisticated obstacle avoidance, and the ability to adapt to changing conditions autonomously. This independence lets drones make decisions in real-time to circumvent obstacles and negotiate challenging environments.

Essential roles that AI and ML can play in improving SRTA include:

Performance optimization: ML algorithms can strategically re-route drones in real-time under variable conditions, such as changing flight durations and adapting to battery constraints. These algorithms are instrumental in determining optimal flight trajectories that ensure mission completion with maximal safety and efficiency.

Real-Time Monitoring and Technical Diagnostics: Applying deep learning models within RTA systems is essential for mitigating unauthorized activities, such as malware, intrusion, or escalating privileges, and assuring that drones operate within established safety and security parameters.

Security Applications: AI and ML can support real-time adaptive rerouting to avoid interception or stop intrusion. They can monitor technical parameters to frustrate tampering or spot cyber-attacks. And they can perform advanced operational health analysis to detect and address security breaches more effectively.

Collision Avoidance: RTA systems are adept at supervising drone control behavior to navigate around obstacles. Innovations in AI and ML can help mitigate collision risks to ensure drones operate safely in diverse environments.

Emergency Landing and Safe Zone Identification: In scenarios requiring emergency landings, AI and ML systems can guide RTA systems to use onboard sensors to identify safe landing zones to significantly reduce risks of ground damage.

Regulatory Compliance: Innovations in AI and ML can help drones stay in compliance with constantly updated regulations, including regulations that change with the drone's location. This is particularly important in urban environments, where drones must operate within strictly defined airspace to avoid conflicts and maintain order. This can help prevent drones from straying into unauthorized areas and maintain legal and safe operation.

Cybersecurity Monitoring: As drones grow more autonomous and more connected, they also become increasingly vulnerable to cyber threats. AI and ML algorithms can help monitor systems and detect unusual behaviors that could indicate cybersecurity breaches that require protective responses.

Technical Failure Monitoring: AI and ML can help extend the scope of RTA to anticipate and detect more types of technical failures. Continuous assessment of a drone›s systems and components can detect malfunctions and initiate corrective actions, keeping the unit on-line and protecting the mission.

Response to Unsafe Operating Conditions: New algorithms for recognizing and responding to unsafe operating conditions, such as adverse weather or congested airspace, can ensure that drones navigate safely through potential hazards. These algorithms can process data about both internal system states and external environmental factors.

Data Handling and Analysis: AI transforms the way drones handle data. Drones equipped with AI-driven processing capabilities can analyze large datasets collected in-flight. This capability improves performance in surveillance, agricultural monitoring, environmental assessments—indeed, in any application that requires tasks at which AI excels, such as image recognition, pattern detection, and predictive analytics.

Swarm Intelligence: Additionally, AI can leverage swarm intelligence to increase coordination among multiple drones. Improved coordination lets the swarm function as a unified entity, capable of executing complex tasks efficiently, precisely, and safely.

NEW ML/AI RTA CHALLENGES

Adopting AI and machine learning in autonomous systems and fleets also introduces new challenges for RTA architects. Engineers will have to ensure that AI-driven drone decisions align with established safety parameters. Developers and quality assurance teams will need to explore how drones can apply AI's advanced decision-making while still allowing RTA elements to provide oversight and step in to maintain operational integrity and safety when needed.

Managing the interaction between RTA and evolving AI algorithms is a complex process. Ensuring that RTA reliably oversees AI decisions requires balancing safety with operational objectives. The unpredictability of AI, particularly of ML-based systems, adds another layer of complexity, requiring RTA systems to identify, manage, and contain these unexpected behaviors effectively.

The statistical nature of these algorithms introduces new trust issues that must be addressed. Furthermore, advanced systems require more processing power than classic RTA algorithms, and compiling and using training data introduces additional technical, regulatory, and ethical considerations. Here are some ways these can show up in RTA analysis and design:

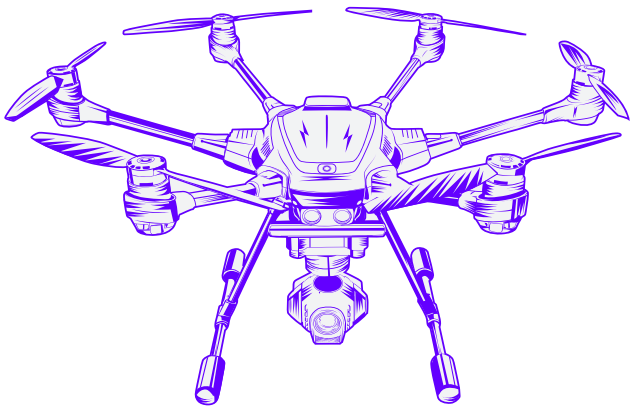
Data quality and real-time processing: AI and ML systems are only as good as the data they are trained on. Inaccurate or biased data can lead to inappropriate AI decisions, taxing the RTA systems' ability to identify and mitigate risks correctly. New frameworks for assessing the quality of AI and ML decision-making can help reduce the impact of faulty decisions based on biased data. Here, the RTA supplements the faulty inputs with real-world observations correlated with new data from multiple sources.

Real-time processing of complex algorithms: Real-time processing and continuous learning necessitate sophisticated data processing capabilities and corresponding adaptations in RTA systems.

Hallucination mitigation: AI models, particularly those built on top of innovations in generative AI and large language models (LLMs), are notorious for hallucinations. New consensus techniques are required to correlate these outputs with live data and robust digital-twin architectures to identify hallucinations and reduce their impact on drone and swarm control systems.

Continuous Learning and Adaptation: Sophisticated adaptive algorithms—capable of online learning and self-improvement—complicate drone AI implementation. New approaches are required to identify issues with compatibility, processing capabilities, and power consumption.

Regulatory alignment: Drone swarms need additional security measures to protect drones from cyber threats and ensure safe operation. And again, it is critical to address regulatory and ethical questions, particularly regarding privacy, accountability, and potential misuse.



SECURE RUN-TIME ASSURANCE: DEFINITIONS AND PRINCIPLES

Secure RTA is a sophisticated framework designed to ensure that autonomous systems, especially drones, operate within their prescribed safety and security boundaries in real time. This concept is particularly crucial for systems that leverage complex, AI-driven processes. Secure RTA employs a layered approach in which an intricate primary system is overseen by a secondary system that is more straightforward and deterministic. This hierarchical supervision is essential for maintaining the security and integrity of these systems during operations amidst external threats and inherent complexities.

Critical Principles of Secure RTA

The effectiveness and implementation of Secure RTA hinge on several foundational principles:

- **Continuous Real-Time Monitoring:** Secure RTA systems are characterized by their relentless, real-time surveillance of operational states, drawing insights from multiple sensors (i.e., sensor fusion) and subsystems. This active monitoring is critical for early anomaly detection, whether due to internal system failures or external threats.

- **Sophisticated Response Mechanisms:** Beyond monitoring, Secure RTA is engineered to respond dynamically to detected threats or deviations. The response spectrum ranges from minor adjustments to complete system overhauls, helping the system adapt to diverse situations and maintain safety and operational continuity.

- **Predictive Analysis:** Using their advanced predictive algorithms, Secure RTA can identify potential risks and operational failures before the threats are fully apparent. These early warnings allow proactive adjustments, averting crises.

- **Balancing Performance and Security:** Crucially, Secure RTA must maintain a delicate balance between robust security and functional efficiency. The goal is to protect against threats while allowing the drone to achieve its operational objectives effectively.

- **Adaptability and Scalability:** In the fast-evolving landscape of drone technology and cybersecurity, the adaptability and scalability of Secure RTA systems are vital. These systems are designed to be flexible, accommodating updates and modifications in response to emerging threats and changing operational requirements.

- **Transparency and Trust:** Secure RTA operates transparently in human-operated or collaborative drone systems to build trust. SRTA provides clear, understandable feedback about the system's status and decision-making processes, which is crucial in scenarios requiring human intervention.

- **Compliance with Regulations and Standards:** Secure RTA systems adhere to relevant safety and security regulations, ensuring responsible and safe drone deployment and operation.

TRUSTED AUTONOMIC SWARM SYSTEMS

Traditional approaches to RTA focused only on improving the safe operation of individual entities. Extending RTA architectures to drone fleets and other types of collaborative AI controllers must allow for emergent behaviors arising from node interactions. Many characteristic problems of managing autonomous drone fleets stem from these unpredictable emergent behaviors and from the complexity of coping with sometimes glitchy communications among tens or hundreds (or even thousands) of moving units. RTA systems are vital for safely and effectively meeting both of these challenges during deployment.

In this larger context, RTA involves real-time monitoring of the swarm's behavior and real-time corrective actions to squelch emergent activity that is unsafe or undesirable. Clearly, this requires a sophisticated and dynamic system that continuously tracks and assesses the activity of each individual unit as well as the fleet as a whole.

University of Bristol researchers have developed the AERoS framework to assess emergent behavior in drone swarms (the term used by the researchers, rather than "fleets"). [16] This framework is domain-independent and applies to many types of swarms. It breaks the assessment into seven stages: safety assurance scoping, safety requirements elicitation, data management, emergent behavior modeling, verification, and model deployment. It operates iteratively and centers on establishing a safety argument pattern based on evidence and artifacts generated at each stage.

However, more work is required to extend this framework to address the security challenges confronting RTA systems for autonomous swarms in real-world environments. For example, the drones might face local issues, like GPS spoofing or spamming, or global issues, like a distributed cyber-attack.

Developers need to devise alternate architectures that are based on local interaction

and allow the swarm to achieve consensus on issues (like insufficient channel resources) that hinder their large-scale coordination. These approaches prioritize the perception-decision-action cycle over reliable information transmission—in other words, the swarm should figure out how to address the underlying problem first, and not fill the airwaves with repeated retransmissions to get their data through. Implementing responses like this depends on fast coordination methods, predictive mechanisms, and robust algorithms to ensure consensus and effective functioning in dynamic environments.

Integrating SRTA in robot fleets is crucial for widespread adoption across industries. Once again, the emphasis in complex swarm operations is on resilience, security, and flexibility. Fleet SRTA should include real-time monitoring, decision-making, and responding to challenges of communication, coordination, and environmental interaction.

TII'S APPROACH TO SRTA

TII's Secure Systems Research Center (SSRC) is exploring how to scale up RTA beyond individual drones to a broader SRTA framework of autonomous systems of systems. This is particularly vital for drone fleets, where the interplay among multiple autonomous units creates a complex network with emergent properties.

Our objective is to embed SRTA in both single entities and the entire collective, guaranteeing that each individual unit and the fleet operate safely and securely. This necessitates a nuanced strategy that incorporates real-time monitoring, decision-making, and adaptive responses to maintain the integrity and reliability of these integrated operations.

By focusing on these systems' dynamic interactions and collective behaviors, SSRC-TII is pioneering a comprehensive RTA framework that concentrates on securing autonomous systems of systems to assure operational effectiveness and resilience in diverse applications. Essential elements of this framework include hierarchical decision-making, safety integrity levels (SIL) assessment, and multiple levels of redundancy.

A Hierarchical Decision-making Approach

We use a hierarchical concept that operates on three levels. The Micro-level analyzes local devices at the edge (Fig. 1). The Meso-level aggregates data from multiple drones in the swarm. And finally, the Macro-level takes a global view of the entire mission, comprising one or more fleets.

Fig 1: Micro-Level Example where edge drone can make its own decisions.

Node (Micro Level)

- Each node has embedded sensors or detectors to identify anomalies (Local SRTA)
- Nodes have localized decision-making capabilities.
- Each node has communication modules to relay decisions or information to other nodes or higher level entities.
- Safety mechanisms like fail-safe controller modes can be embedded at this level.
- Safety levels (SIL) can be assigned to decisions made at this level to signify their importance.

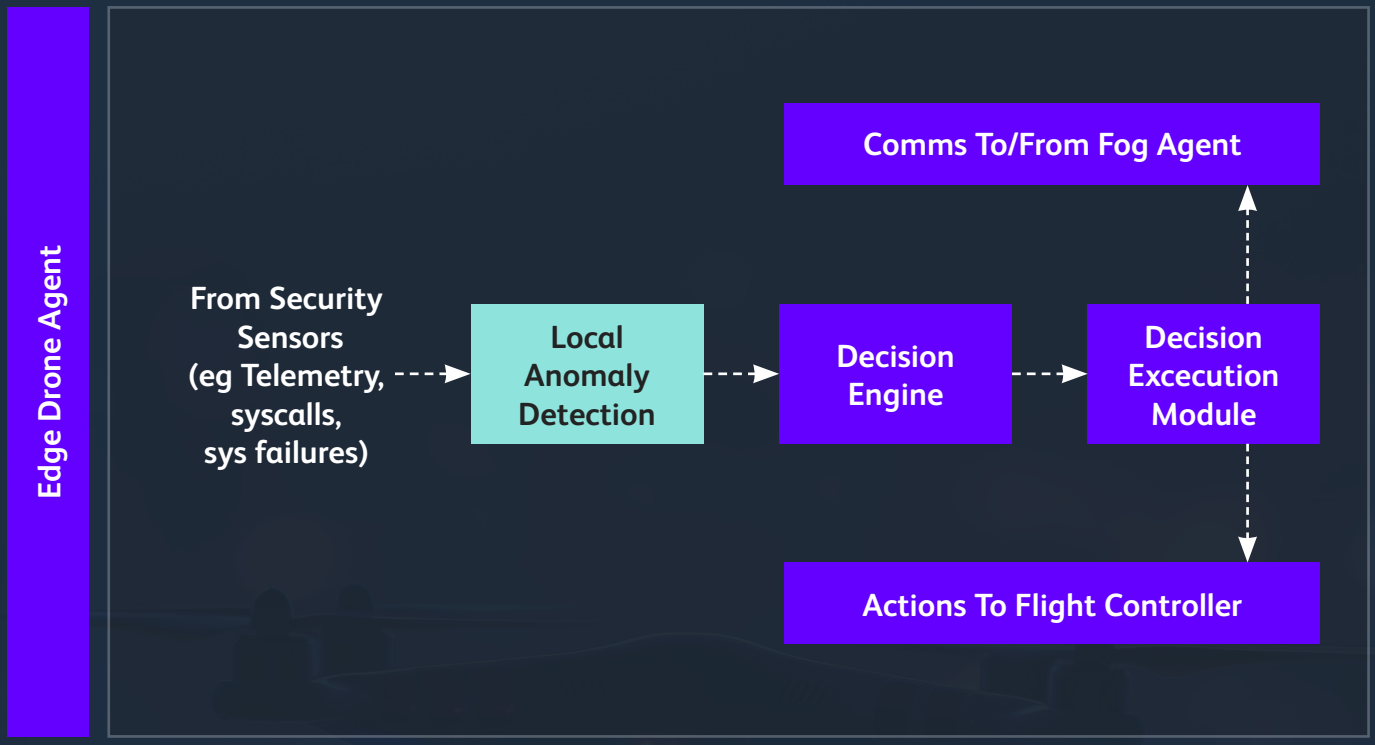


Fig 2: General diagram of the Edge Drone SRTA Agent.

Micro-Level

At the micro-level (Fig. 2), we have implemented a comprehensive monitoring strategy, using multiple probes or sensors to track a wide range of system events, including security incidents, physical sensor telemetry, operating-system activities, and communication exchanges, among others. These sensors interact with diverse machine-learning algorithms, enabling real-time data monitoring.

Each node in this network is equipped with embedded sensors for anomaly detection. Each node effectively serves as a localized SRTA system and boasts localized decision-making capabilities. Communication modules allow nodes to relay decisions or pertinent information to other nodes or higher-level entities within the system. Mechanisms such as fail-safe controller modes are integrated at this micro level to bolster safety, to add redundancy to the system.

When an anomaly is detected, the SRTA promptly reports it to a central decision engine. The decision engine, running locally on each drone, is equipped with advanced error handling and self-healing capabilities, ensuring uninterrupted operation by automatically detecting, isolating, and correcting system faults. It employs encrypted communication protocols, robust access control, and continuous security monitoring, safeguarding against unauthorized access and maintaining the highest standards of data security

The system assigns safety integrity levels (SILs) to the decisions made, reflecting their criticality and their importance in maintaining overall system integrity and security. This multi-faceted approach ensures a robust and responsive system capable of identifying and reacting to potential issues swiftly and efficiently.

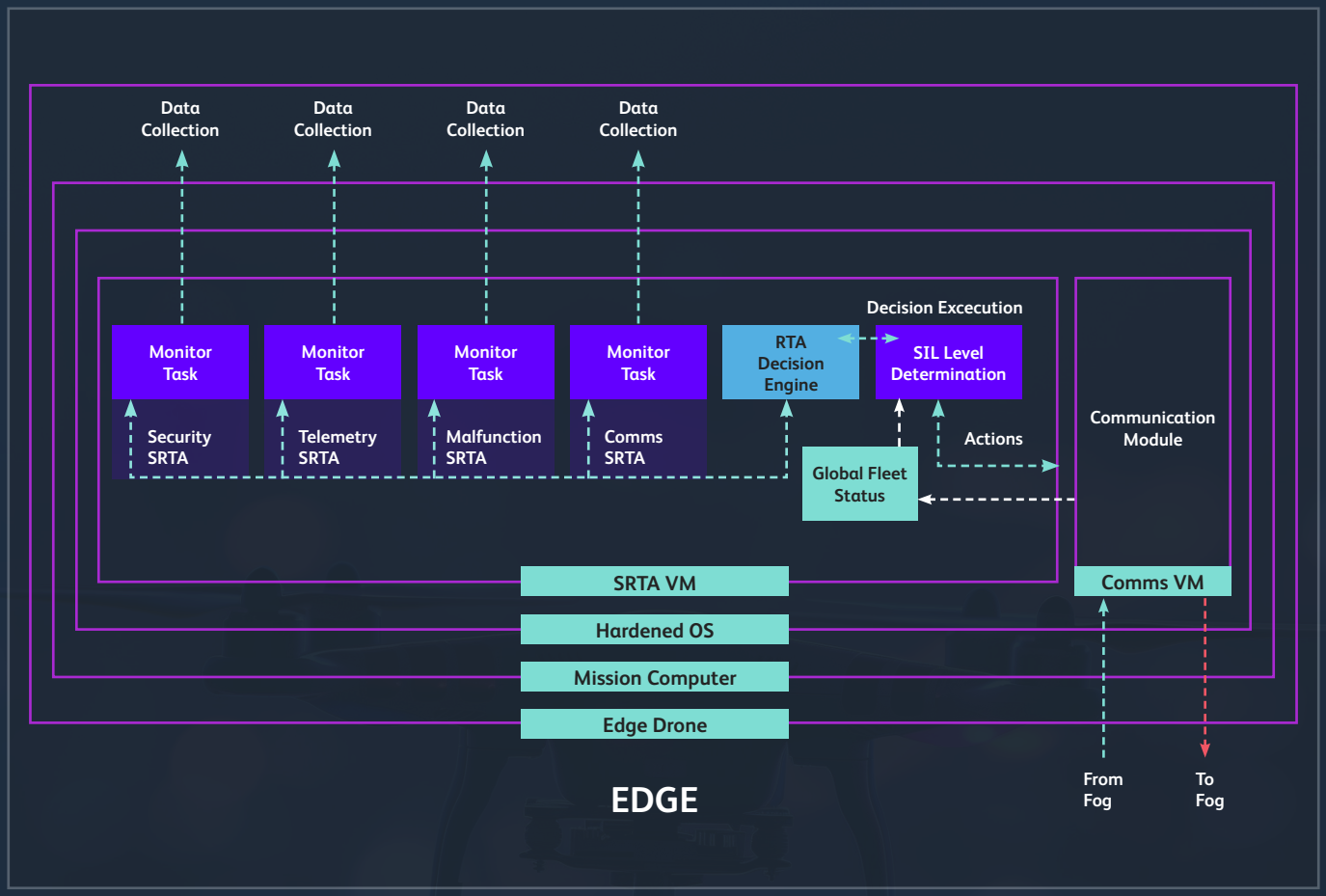


Fig 3: Implementation of the Edge Drone SRTA Agent.

The SRTA system uses an integrated hierarchical architecture (Fig. 3), which focuses on aggregating and processing information from a single drone's perspective. This architecture encapsulates a multi-layered decision-making process, ensuring robust and responsive operation within the drone swarm ecosystem.

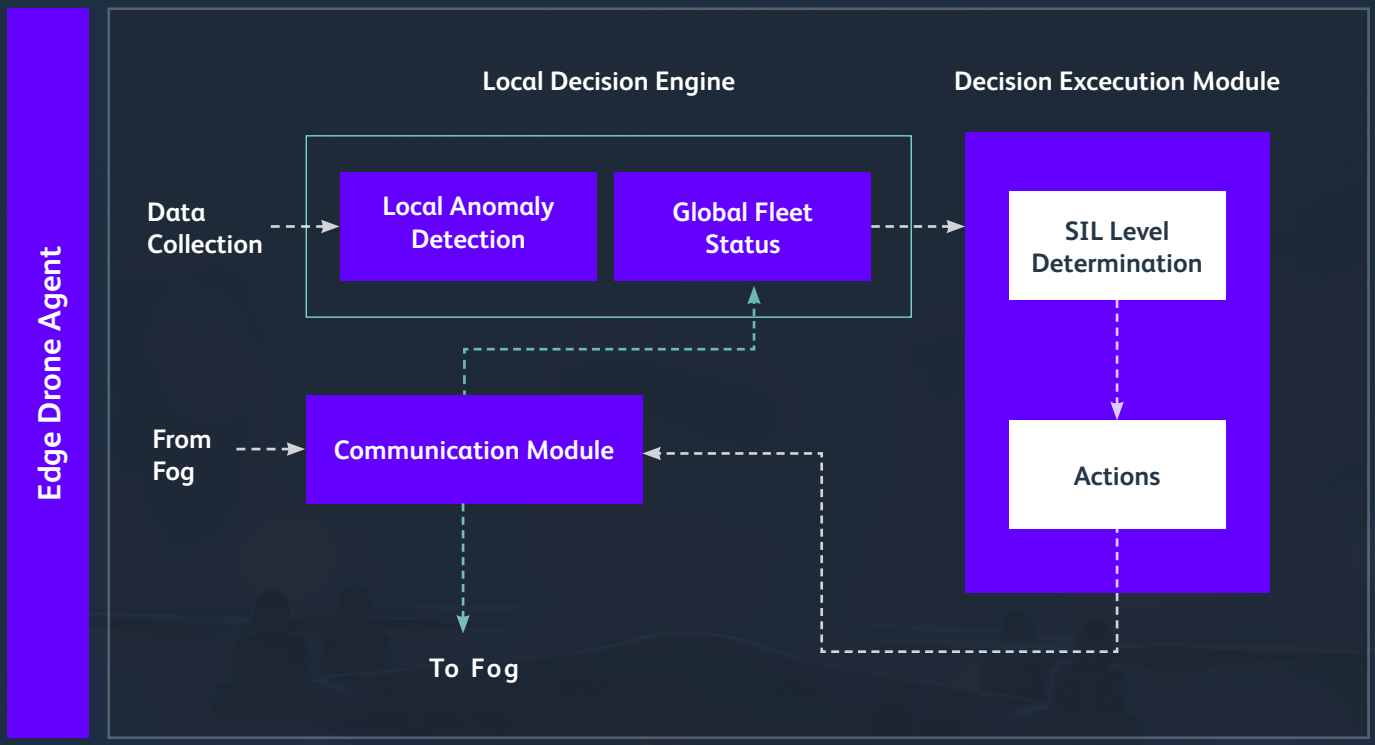


Fig 4: Detailed view of the Edge Drone SRTA Agent components.

Figure 4 shows the view of the Edge Drone SRTA Agent. This system's core is the local decision engine, bifurcated into two essential submodules: the Anomaly Detection System and the Global Fleet Status module. The Anomaly Detection System continuously monitors the drone's local environment. Its array of sensors vigilantly scans the system for any irregularities or deviations. Concurrently, the Global Fleet Status module assimilates information received from the swarm's other drones via a dedicated communication module. This dual-input approach ensures a comprehensive overview that covers both internal system status and external fleet dynamics.

The heart of the decision-making process is the Decision Execution Module. This is the epicenter of the system's intelligence and is responsible for evaluating the gathered data to determine the appropriate Safety Integrity Level (SIL) and determine the corresponding action. Here, critical assessments are made, weighing the severity of detected anomalies against predefined safety parameters to tailor a response.

Once the system reaches a decision, it synchronizes this information with its peers—or,

if necessary, escalates the issue to a higher operational level, such as the fog drone. Again, the advanced communication module facilitates this communication, ensuring seamless and efficient exchange of vital information. This hierarchical-yet-interconnected architecture underlines the system's capability to make autonomous decisions at a local level and integrate these decisions into the broader swarm context, maintaining optimal operation and safety throughout the network.

Meso-level

At the meso-level (Figs. 5 & 6), our system deploys a sophisticated aggregation process to compile information from multiple drones, thereby building a comprehensive view of fleet dynamics. Each drone in the fleet functions as an individual agent, yet they collectively form a complex network for developing collaborative or collective behaviors. This synergy is pivotal in the fleet's functionality, particularly when the collection must adapt to varying mission conditions. Depending on the mission's specific requirements, this aggregated information can either be directed to a centralized node—such as a fog drone (a more capable unit that consolidates and integrates input from the drones in its immediate area and combines it with its own direct sensing of the area, topographic data, and other information to form a consolidated picture and plan of its local area—cutting through the "fog of war"—and reporting it up the chain of control). Or the data can be pushed out for more distributed processing. In the latter scenario, all drones within the swarm collaborate to reach a consensus before any decision is executed, producing a cohesive and unified response.

At this meso-level, the system is designed to oversee the swarm's dynamics and make informed decisions based on the nodes' collective behavior and interactions. The meso-level receives and processes information relayed from individual nodes and then synthesizes the data to form a broader understanding of the fleet's state and activities. Furthermore, this system-level is equipped to prioritize decisions based on their criticality. SIL plays a crucial role here, providing a standardized framework for assessing and

categorizing the urgency and importance of each candidate decision. This hierarchical decision-making approach increases confidence that critical decisions are given precedence—especially when they might impact operational safety and efficacy.

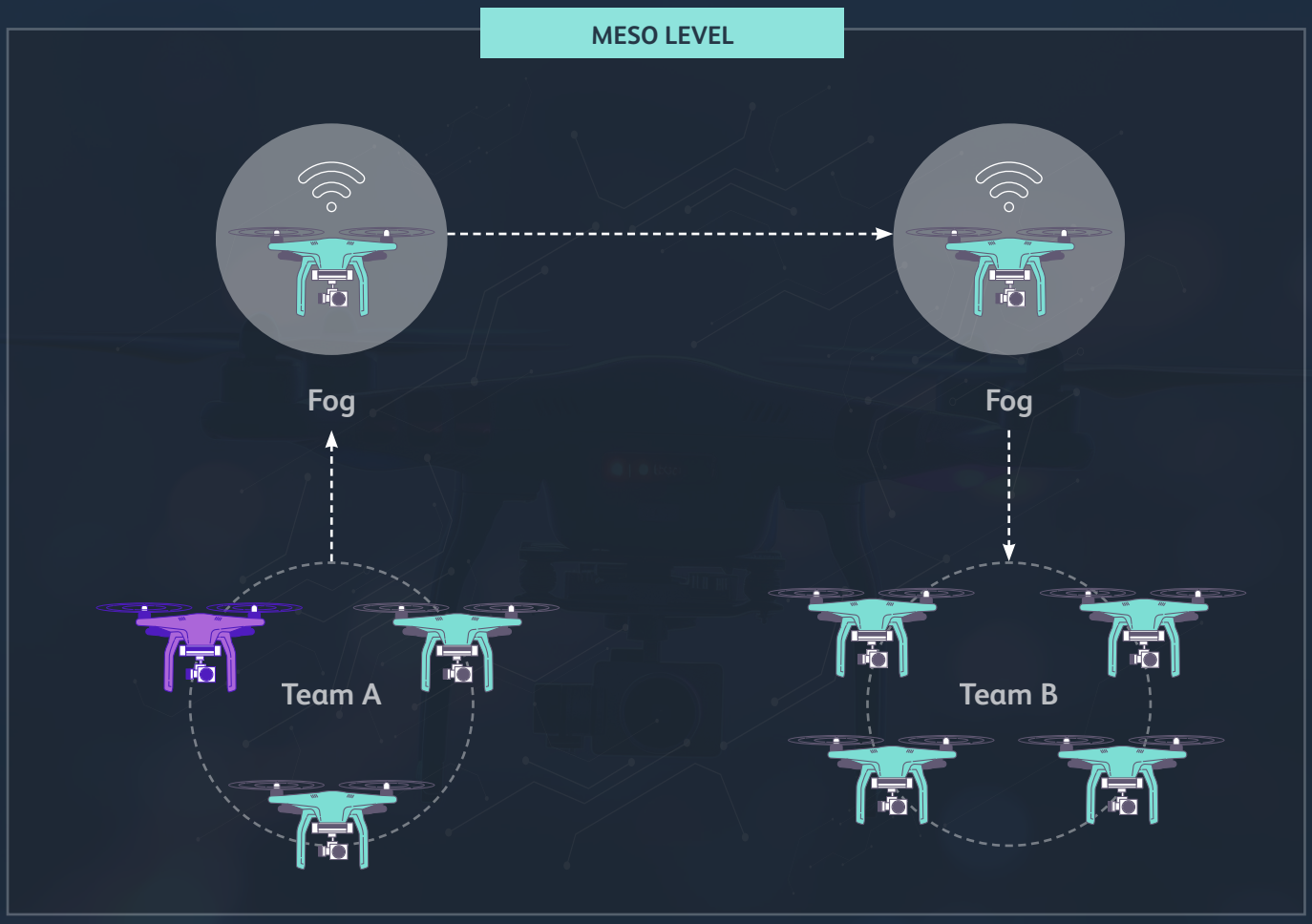


Fig 5: Meso Level example where a Fog Drone collects the fleet information.

Fog Drone (Meso Level):

- Overlooks the swarm dynamics and can make decisions based on the collective behavior of nodes.
- Receives information from nodes and processes it.
- Can prioritize decisions based on their criticality (SIL levels can help here).
- Communicates with the GCS for significant decisions or when a higher-level perspective is required.

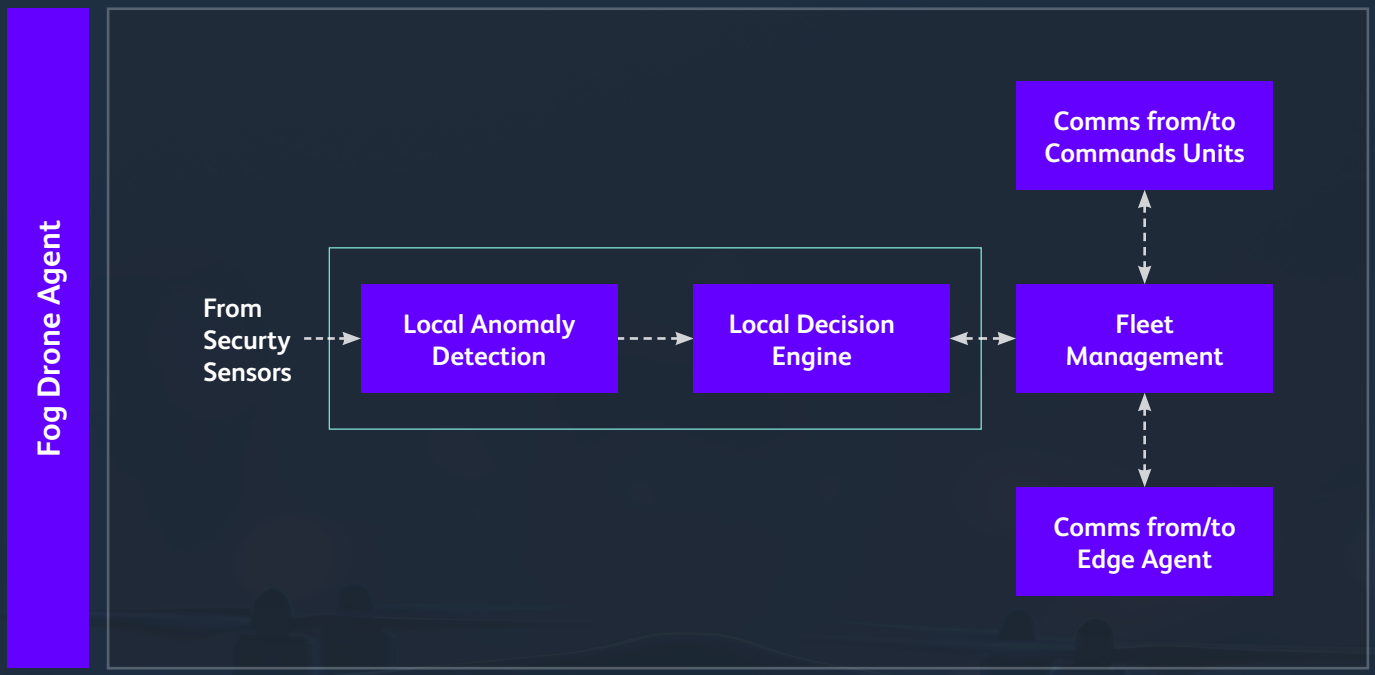


Fig 6: General Diagram of the Fog Drone Agent.

Macro-level

At the macro-level of our system architecture (Figs. 7 & 8), the Cloud or Ground Control Station (GCS) assumes ultimate decision-making authority, particularly in mission-critical scenarios. This is the highest echelon of operational oversight, with the power to override decisions made by intermediary entities like fog drones or individual nodes—especially in circumstances deemed non-recoverable. In such critical situations, the Cloud or GCS level is not just a passive information recipient; rather, it is a proactive decision-maker equipped to assess, strategize, and direct actions to mitigate risks and preserve mission integrity.

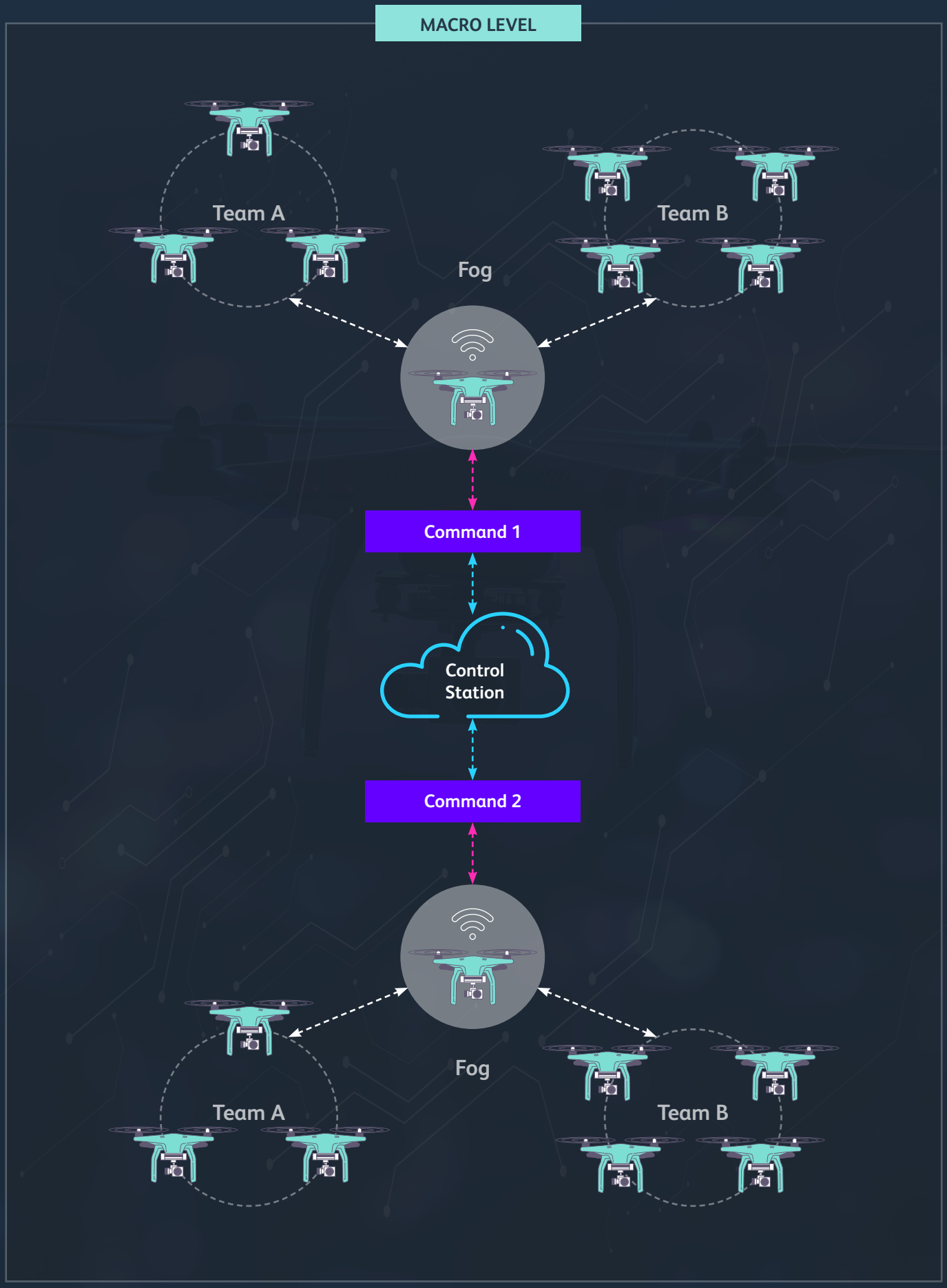


Fig 7: Macro Level overall integration and control of fleet-of-fleetss.

Cloud or GCS Level (Macro Level)

- Ultimate decision-making authority, especially for mission-critical scenarios.
- Can override decisions made by the fog drone or nodes in case of non-recoverable scenarios.
- Uses SIL or safety levels to strategize the best course of action.

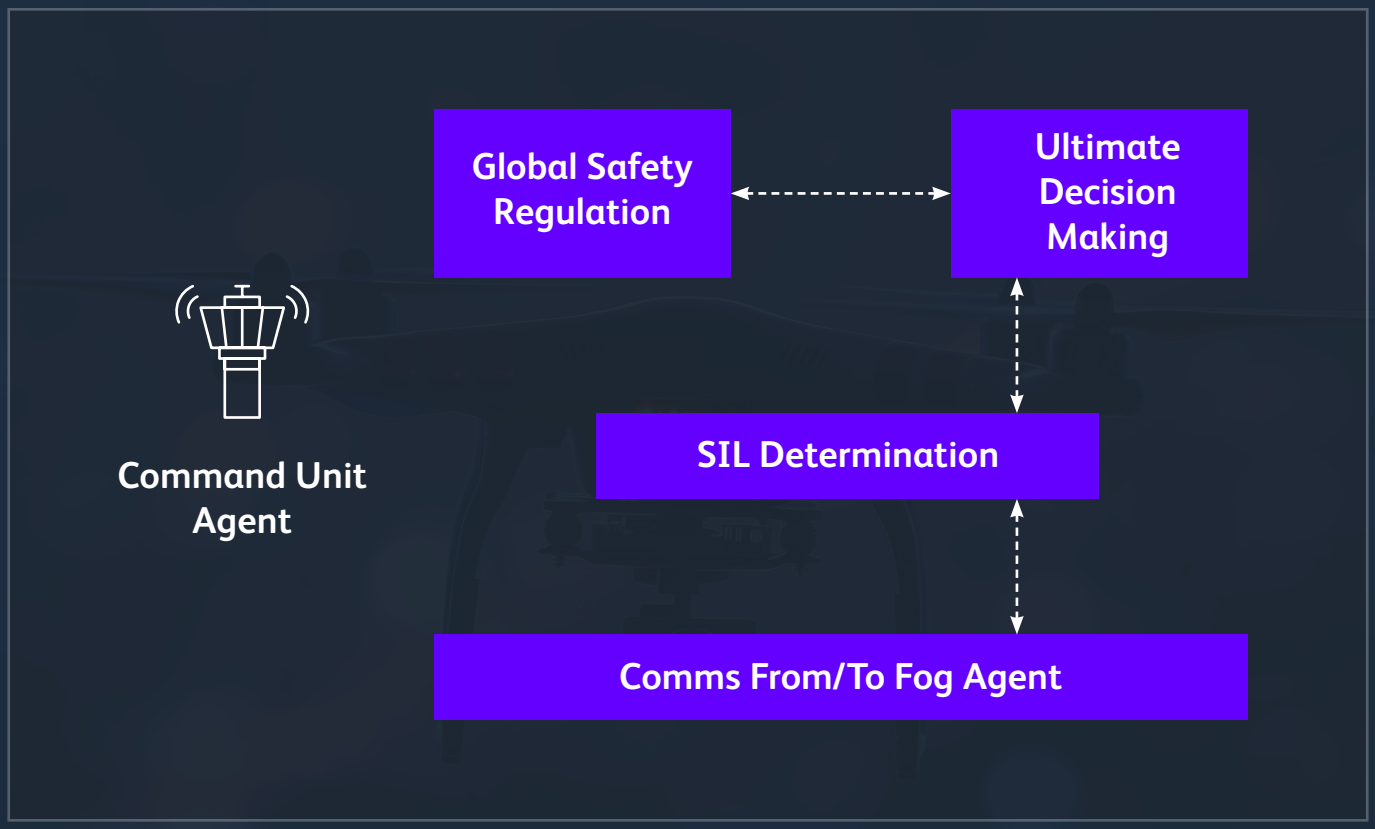


Fig 8: Flow Diagram of the Command Unit Agent.

Safety Integrity Levels (SILs)

Safety Integrity Levels (SILs) are integral at this macro level, where they serve as guidelines for prioritizing actions based on their criticality and potential impacts on the operation. By employing SILs, the Cloud or GCS can effectively game out the best course of action to yield responses that are as timely as possible and as closely aligned as they can be with the severity and urgency of the situation at hand. This hierarchical and safety-conscious approach underpins the system's robustness so that even in the most challenging scenarios, decision-making is guided by a well-defined, systematic, and safety-oriented framework. Therefore, the Cloud or GCS level stands as the backbone of the decision-

making process, providing the strategic oversight crucial for successfully executing and completing complex missions.

Our system architecture adopts a comprehensive and hierarchical approach for evaluating anomalies and making decisions—guided by Safety Integrity Levels assigned to reflect the potential impact of each detected anomaly or decision. This structured approach means that responses are proportionate to the severity of each situation.

At the foundational level, SIL-A is assigned to minor anomalies that do not threaten the mission (e.g., a minor sensor calibration issue in one drone). While these anomalies require attention, they do not require urgent or drastic responses. Moving up the scale, SIL-B designates moderate anomalies that might affect the mission, but the impact is typically recoverable (such as a temporary communication disruption between a single drone and the swarm). In more severe scenarios, SIL-C is assigned to significant anomalies, those that significantly impact the mission (e.g., a portion of the swarm experiencing GPS signal jamming). These issues, while grave, are still potentially recoverable, though they may demand substantial corrective action. The most critical level, SIL-D, is reserved for anomalies that pose a dire threat to the entire mission and demand immediate and decisive action (e.g., a cyber-attack targeting the swarm's whole control system).

The influence of SILs on decision-making is profound. Higher SIL ratings trigger heightened attention and a more immediate response from the relevant entities in the system's hierarchy. For instance, a decision or anomaly classified as SIL-D at the node level would promptly escalate to the GCS, bypassing intermediate levels to ensure a rapid and effective response. This tiered SIL-based approach to anomaly classification and decision-making is instrumental in maintaining operational integrity and safety, while also allowing for calibrated responses proportionate to the severity of the situation.

Redundancy

Incorporating redundancy into the drone RTA framework is critical to increasing reliability and safety, especially in scenarios where system integrity is paramount. For example, a system might require a secondary flight controller to serve as a backup if the primary controller is compromised. Early detection of a primary controller malfunction or security breach allows a measured transition to the secondary controller, sustaining continuous, safe operation, and mitigating the risk of catastrophic failure.

Here, the secondary flight controller is not merely a passive component waiting to be activated; it continuously monitors the primary controller's status and the drone system's overall health. In real-time, the backup evaluates the primary's data and decisions and stands ready to intervene if it detects anomalies or deviations. Such redundancies are particularly crucial when drones operate in challenging or hazardous environments, or when the missions are critical – as in search-and-rescue missions, infrastructure inspection, or surveillance.

Moreover, this redundancy aligns with the broader objectives of RTA, which emphasize operational safety and security throughout the mission. Having redundant systems in place bolsters the RTA framework, safeguarding the drone's operational integrity and its ability to perform its designated tasks effectively and safely, even in the face of unexpected failures or external threats. This design philosophy instills confidence in operators who know that robust defenses are in place to handle unforeseen events. SSRC is focusing on redundancy as a key strategy to safeguard against failures and recover from them.

THE FUTURE

The future of runtime assurance in drone technology will be characterized by a blend of technological innovation and an increasing focus on societal and ethical integration. By continuously evolving to incorporate advanced AI and ML and adapting to complex operational environments, RTA systems are poised to play a crucial role in the safe and efficient operation of autonomous drones in every sector where they find employment. Addressing the great challenges—of scalability, regulatory compliance, and the balance between safety and efficiency—will be essential in realizing this vision.

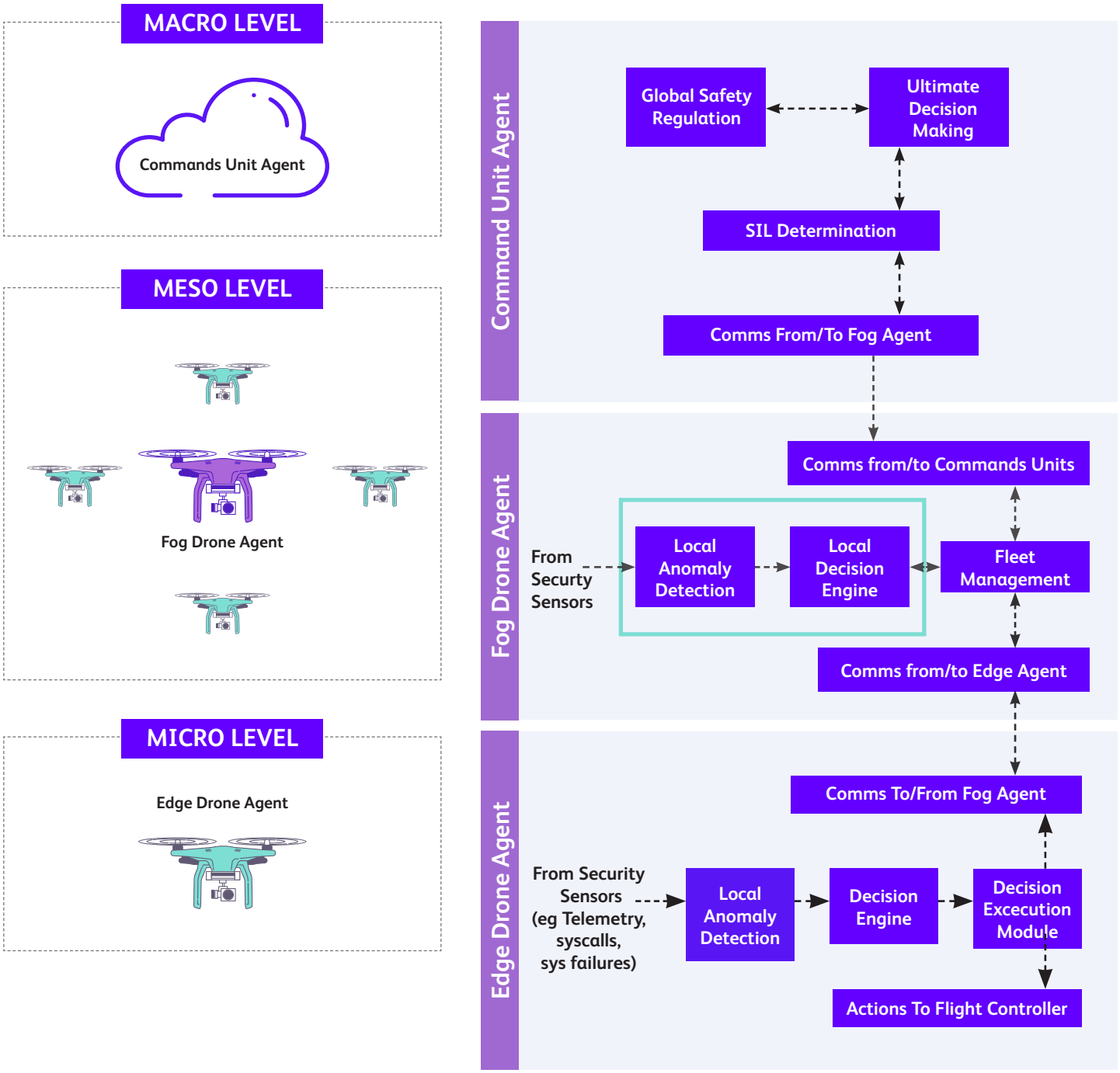
Growing AI and ML capabilities will boost RTA systems' predictive capabilities, producing more accurate and timely responses to potential threats and operational anomalies. A key challenge will be maintaining the delicate balance between operational efficiency and safety. Future RTA systems will need to guarantee that safety measures do not impede the drones' functional efficiency and that functional efficiency does not jeopardize safety.

As drones and embedded controllers become ever more autonomous, RTA systems must adapt to managing these systems in increasingly complex environments. This includes urban air mobility and intricate logistics operations, where drones could be pivotal in passenger transport and supply-chain management. With the growing use of drone swarms, RTA systems must scale effectively to manage the increasing complexities of large-scale, coordinated operations.

The long-term goal is to develop RTA systems that enable drones to operate autonomously in the widest possible variety of environments. This level of autonomy necessitates RTA systems that can handle complex decision-making processes while ensuring the highest safety and security standards.

Future RTA systems will also need to address ethical, regulatory, and technical challenges as drones become more integrated into societal functions. As drone technology becomes a global, everyday reality, standardizing RTA systems and ensuring compliance with international regulations will become crucial.

RESILIENCE ACROSS SYSTEMS



REFERENCES

1. "ATSB Investigates after Hundreds of Drones Plunge into Yarra River." https://australianaviation.com.au/07/2023/atsb-investigates-after-hundreds-of-drones-plunge-into-yarra-river/
2. Lee et al., "Runtime Assurance Based On Formal Specifications." https://www.researchgate.net/publication/46176361_Runtime_Assurance_Based_On_Formal_Specifications
3. L. Sha, «Using simplicity to control complexity,» IEEE Software, vol. 18, no. 4, pp. 28-20, Jul. 2001, doi: 10.1109/MS.2001.936213.
4. M. Clark et al., «A Study on Run Time Assurance for Complex Cyber Physical Systems,» Defense Technical Information Center, Fort Belvoir, VA, Apr. 2013. doi: 10.21236/ADA585474.
5. J. G. Fuller, «Run-Time Assurance: A Rising Technology,» in 2020 AIAA/IEEE 39th Digital Avionics Systems Conference (DASC), Oct. 2020, pp. 9-1. doi: 10.1109/DASC50938.2020.9256425.
6. Y. Peng, G. Tan, and H. Si, «RTA-IR: A runtime assurance framework for behavior planning based on imitation learning and responsibility-sensitive safety model,» Expert Systems with Applications, vol. 232, p. 120824, Dec. 2023, doi: 10.1016/j.eswa.2023.120824.
7. «Trusted Autonomy | GE Research.» Accessed: Dec. 2023 ,08. [Online]. Available: https://www.ge.com/research/initiative/trusted-autonomy
8. E. C. Research, «An Overview of Draft UL 4600: ‹Standard for Safety for the Evaluation of Autonomous Products,›» Medium. Accessed: Dec. 2023 ,08. [Online]. Available: https://edgecaseresearch.medium.com/an-overview-of-draft-ul-4600-standard-for-safety-for-the-evaluation-of-autonomous-products-a50083762591
9. K. Hobbs, M. Mote, M. Abate, S. Coogan, and E. Feron, «Run Time Assurance for Safety-Critical Systems: An Introduction to Safety Filtering Approaches for Complex Control Systems,» IEEE Control Syst., vol. 43, no. 2, pp. 65-28, Apr. 2023,

doi: 10.1109/MCS.2023.3234380.

10. «Advancing autonomy in Australia | BAE Systems.» Accessed: Dec. 2023 ,08. [Online]. Available: https://www.baesystems.com/en-aus/what-we-do/autonomy
11. «Bringing a Pilotless Future and Autonomous Delivery One Step Closer to Reality | Auterion.» Accessed: Dec. 2023 ,08. [Online]. Available: https://auterion.com/bringing-a-pilotless-future-and-autonomous-delivery-one-step-closer-to-reality/
12. "Introducing the Ghaf Platform: An Innovative Solution to Zero Trust Architecture." [Online]. Available: https://engineeringresources.tradepub.com/free/w_tecm10/
13. "Zero Trust Architecture in Combatting Wireless Network Attacks." [Online]. Available: https://engineeringresources.tradepub.com/free/w_tecm07/
14. "Building a Zero Trust Security Model for Autonomous Systems - IEEE Spectrum." [Online]. Available: https://spectrum.ieee.org/zero-trust-security-autonomous-systems
15. "Ready for the Ultimate Security Solution? Create a Safer World with Zero-Trust Autonomic Swarm Security." [Online]. Available: https://engineeringresources.tradepub.com/free/w_tecm17/
16. D. B. Abeywickrama et al., «AERoS: Assurance of Emergent Behavior in Autonomous Robotic Swarms,» vol. 2023 ,14182, pp. 354-341. doi: 28_0-40953-031-3-978/10.1007.

Technology Innovation Institute LLC
P.O. Box 9639
Abu Dhabi, UAE

tii.ae