Supporting Mission Assurance and Securing Autonomous Systems through Cyber-worthiness Principles

MILCIS 2023

Iain Dickson

14th November 2023

SCOPE

- Definition of Terms
  - Information Technology vs Operational Technology
  - Autonomy and its maturity stages
  - Cyber Security and its maturity stages
    - Cyberworthiness
  - Mission Assurance
    - Relationship between Cyberworthiness and *worthiness in achievening Mission Assurance
- Case study comparison of autonomous and non-autonomous systems
- Integrating Safety and Security through a risk management process
  - Implementation of Mitigations within the process.
- Why integrate safety and security?

BIOGRAPHIES

Iain Dickson

Full Spectrum Cyber Practice Lead

Iain is the Full Spectrum Cyber Practice Lead for Leidos Australia, providing oversight and support to all of Leidos' Australian programs for cyber security including its military platform work. Iain has previously worked as a Cyber Research Engineer and an Assistant Director for Cyber Threat Intelligence within the Federal Government. Iain is one of the founders of ComfyCon AU, a virtual conference founded in response to the cancellation of cyber security conferences due to the COVID-19 pandemic.

Kate Tollenaar

Trusted Data and AI/ML Practice Lead

Kate is the Trusted Data and Artificial Intelligence/Machine Learning Practice Lead at Leidos Australia, assisting with the delivery of trustworthy, resilient and secure data solutions in Leidos' programs. Prior to this role, Kate was in the Australian Army as a Signals Officer for 23 years. Kate holds a Bachelor of Arts (Honours) from the University of New South Wales, a Masters in Strategy and Security and a Masters of Applied Cybernetics.

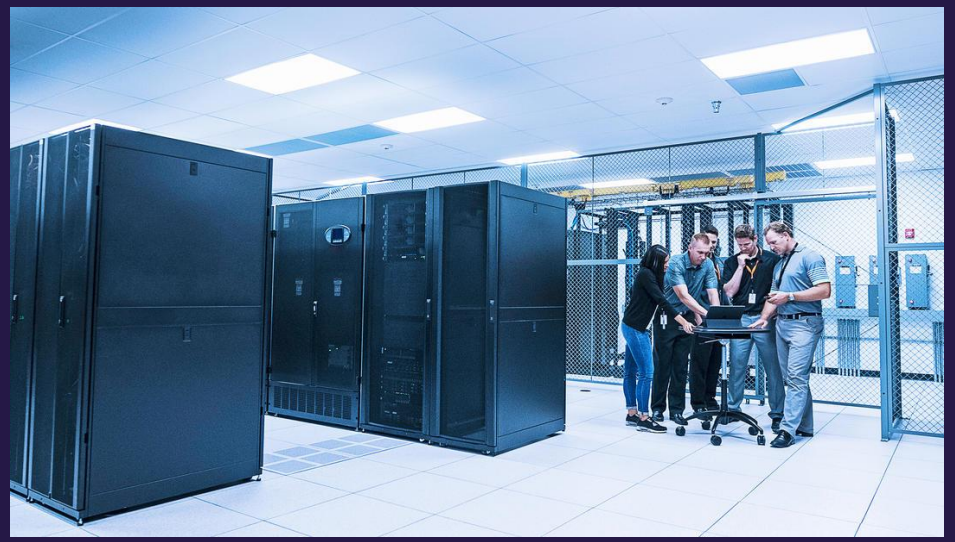
The beginning of wisdom is the definition of terms

SOCRATES

INFORMATION VS. OPERATIONAL TECHNOLOGY

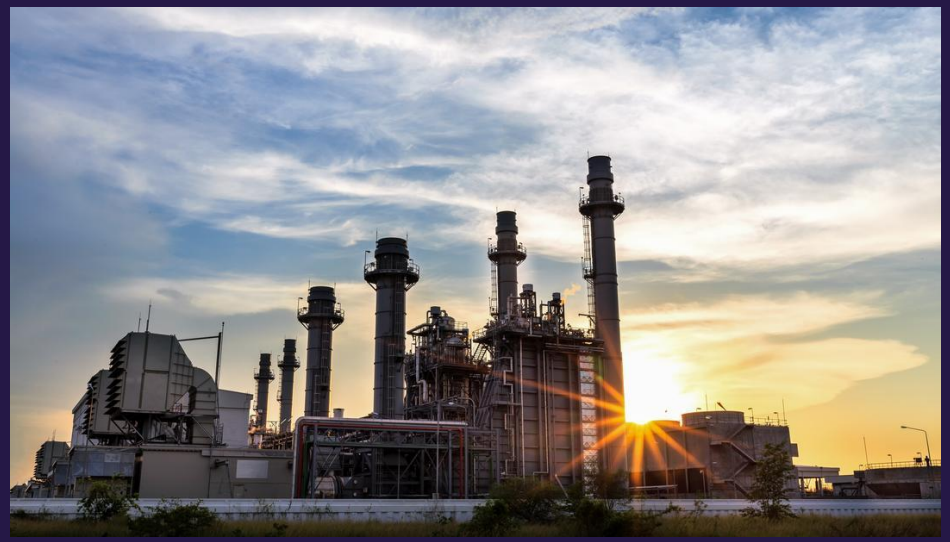
Computing and/or communications hardware and/or software components and related resources that can collect, store, process, maintain, share, transmit, or dispose of data.

Source: NIST Information Technology



Programmable systems or devices that interact with the physical environment (or manage devices that interact with the physical environment).

Source: NIST Operational Technology



AUTONOMY AND ITS MATURITY STAGES

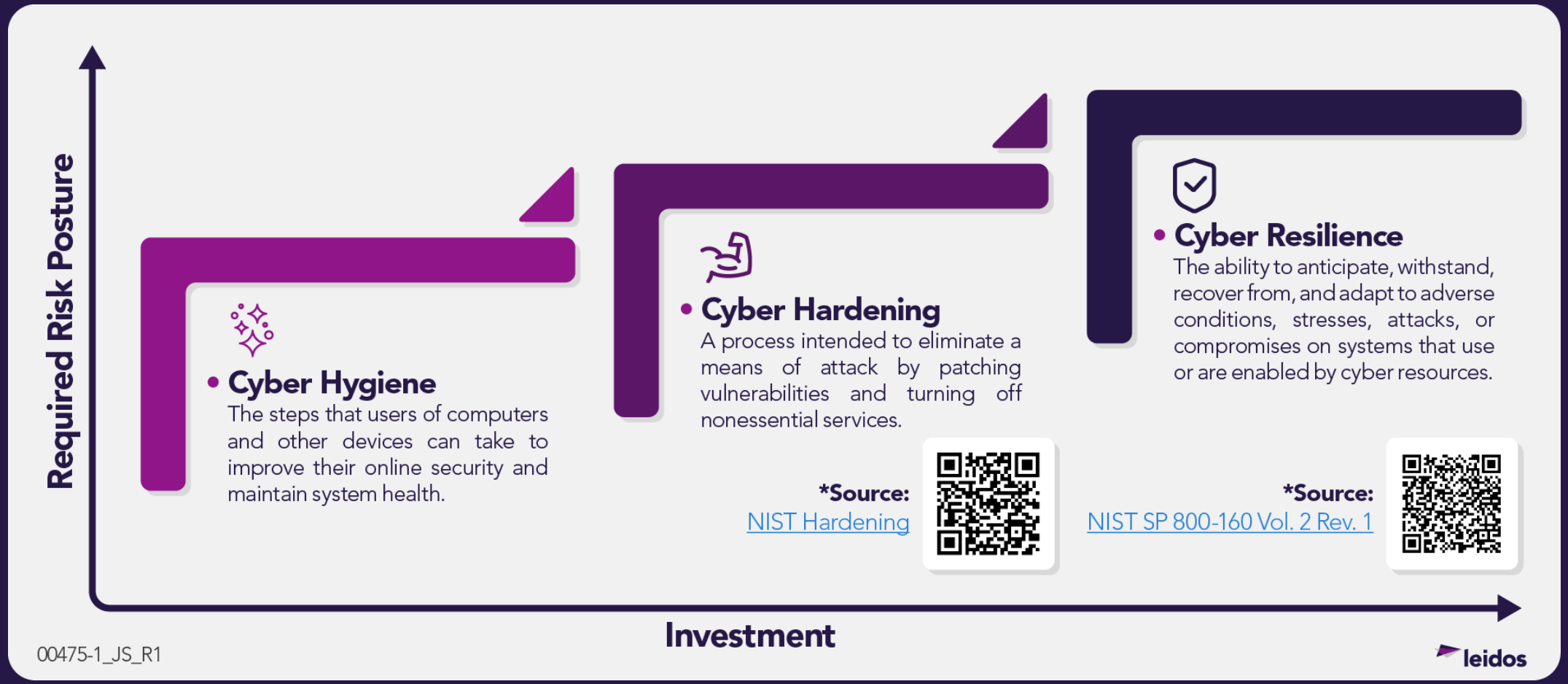
- Autonomy maturity is the level of human input required to operate the system and the level of technical ability of the system.

- In general, higher levels of platform autonomy will require more sophisticated AI technology to collect, analyse and reason with and achieve assigned goals.
- The convergence of IT and OT in autonomous systems offers levels of autonomy for human vs machine control.

Level of Autonomy	Human Role	Agent Role	Autonomous Capability
Remotely Operated	Full human control	Human dependant	**Actuate platform sub-systems:** Move steering wheel, push brake
Automatic	Monitor-intervene (human in the loop)	Self-operating	**Complete actions:** Pre-programmed movements in sub-system
Autonomic	Supervise-task (human on the loop)	Self-managing	**Achieve tasks:** Interpret information to dynamically achieve individual tasks (autonomous sub-systems)
Autonomous	Collaborate-rely (human starts the loop)	Self-governing	**Achieve goals:** Plan and achieve multiple tasks to reach final objective

Source: Australian Army Robotics and Autonomous Systems Strategy version 2

CYBER SECURITY AND ITS MATURITY STAGES



CYBERWORTHINESS

An assessment of the cyber resilience of a system from cyber attack within a given mission context. This includes its ability to:

- Operate while under degraded, denied, intermittent and low bandwidth (DDIL) scenarios.
- Continue to operate whilst under attack and provide mission value.
- Deliver the mission outcome

- Cyberworthiness controls are domain (sea, air, land) agnostic, but context (urban, rural, ocean) dependent.
  - Using insecure WiFi in the middle of the ocean, whilst bad, is probably okay if it enables the mission.
  - Using insecure WiFi in the middle of an urban environment, is probably a bad idea.
- It's important to include that security controls need to be continuously considered, updated and implemented given the changing contexts in which the system is being used.
  - Moving a system that is generally used in an ocean environment, to an urban environment or a harbor environment.

MISSION ASSURANCE

- **Mission Assurance**
  A process to protect or ensure the continued function and resilience of capabilities and assets—including personnel, equipment, facilities, networks, information and information systems, infrastructure, and supply chains—critical to the execution of organizational mission-essential functions in any operating environment or condition.

***Source:**
NIST SP 800-160 Vol. 2 Rev. 1

00475-2_JS_R1

leidos

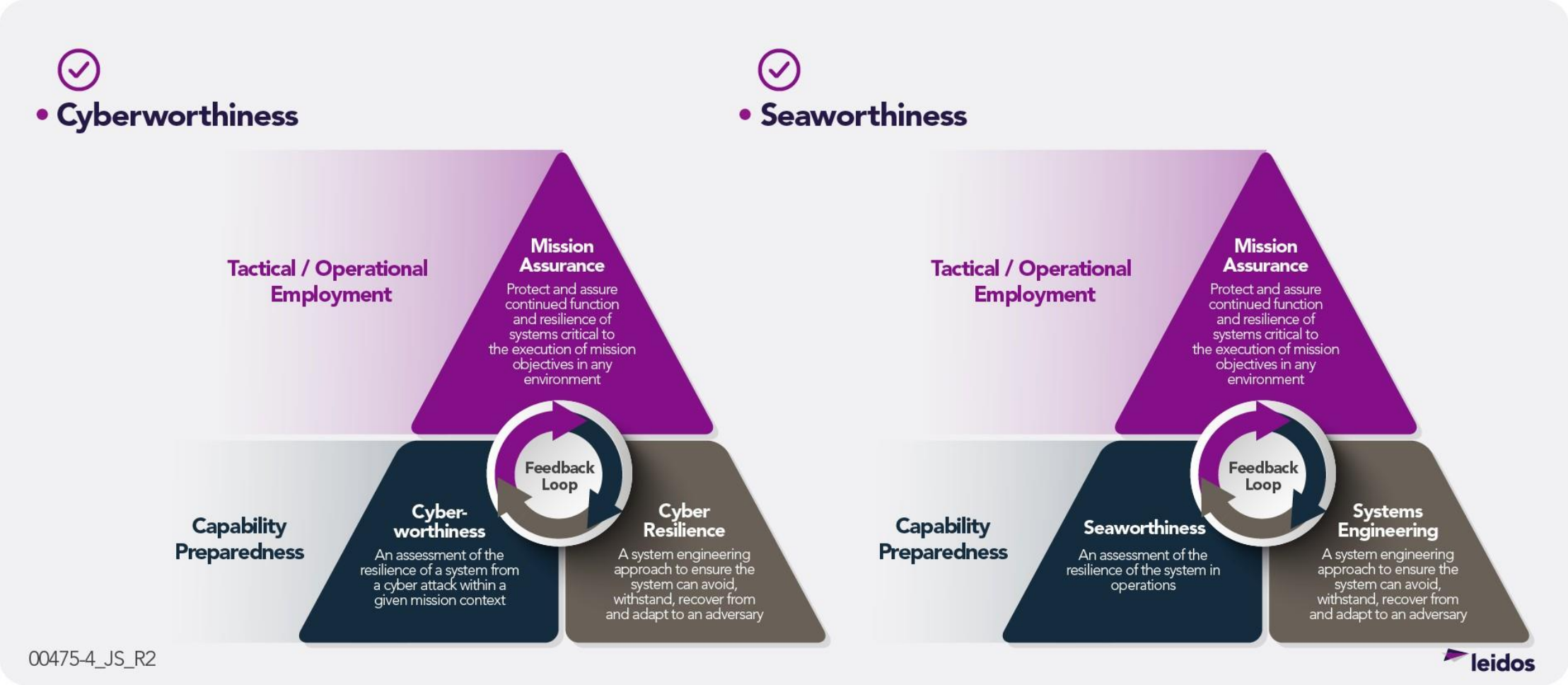
CYBERWORTHINESS AND MISSION ASSURANCE

- Cyber Worthiness and Mission Assurance

00475-3_JS_R1

leidos

RELATIONSHIP BETWEEN CYBERWORTHINESS AND MISSION ASSURANCE



The Three Laws of Robotics:
1. A robot may not injure a human being, or through inaction allow a human being to come to harm…

ISAAC ASIMOV 1942

CASE STUDY – AUTONOMY VS NON AUTONOMY

- Variant 1 – Operated:
  - A non-autonomous aircraft is traversing above Melbourne Airport – an extremely busy airspace with many aircraft containing civilians. This aircraft is controlled by a SQEP crew, who evaluate information provided by systems and sensors to change the physical effects of the aircraft (speed, direction, etc…) and make sure they can reach their final location.
- Variant 2 - Autonomous:
  - An autonomous aircraft is traversing above Melbourne Airport – an extremely busy airspace with many aircraft containing civilians. This aircraft is controlled by an autonomous system software, which evaluates information provided by systems and sensors and enacts actions on the environment to reach the systems final location.
- A threat actor has been able to infiltrate the software systems of each aircraft in order to adjust the sensors so that all detected objects are 45 degrees right of where they are in the real world. This is a traditional integrity attack on a system.

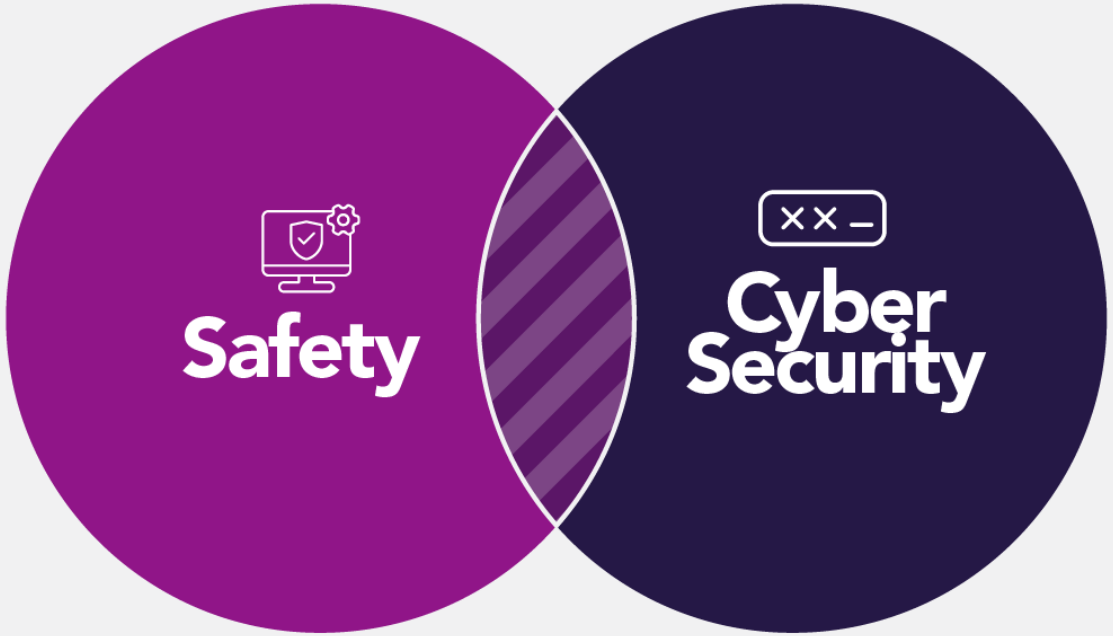
NON-AUTONOMOUS SYSTEMS RESPONSE

1. Humans are involved in the running of the aircraft and are trained and certified in using it.
2. Sensor provides incorrect data to the humans for them to make a decision.
3. The humans identify there is a fault in the aircraft as they see an object where the sensors say there should not be one.
4. The humans take remedial action, either identifying the offset or ignoring that sensor data and continuing the mission.

While the security of the system is important in this case, the immediate mission can continue without the input from the sensor.

CYBER SECURITY AND SAFETY WITHIN NON AUTONOMOUS SYSTEM

- Cyber Security and Safety Within a Non-Autonomous System



00475-5_JS_R1

leidos

AUTONOMOUS SYSTEMS RESPONSE

1. The autonomous software system runs within the aircraft and the system has been verified and validated within a given set of conditions.
2. Sensor provides incorrect data to the autonomous software system for it to make a decision.
3. The autonomous software system is driven by a) what it sees (sensor data) and b) its guard rails and makes a decision not to hit an aircraft at 45 degrees right by turning 45 degrees left.
4. The autonomous aircraft hits another aircraft in the airspace.

In this situation, there is a direct relationship between the safety of the platform, and the security of the platform.

CYBER SECURITY AND SAFETY WITHIN AUTONOMOUS SYSTEM

- Cyber Security and Safety Within an Autonomous System

00475-6_JS_R1

leidos

OTHER POTENTIAL EFFECTS

The following are examples of effects that could be induced within an autonomous system through cyber attack

- Turning off vital systems such as transponders, engines or communications.
- Redirect the autonomous system to another location, including to an adversary or even back to the control headquarters.
- Use the autonomous system to achieve an alternate effect or diversion.
- Disposal of payload or cargo.

Risk does not define a current problem or a future certainty, but rather the potential for future harm.

Risk management is not about future decisions, but about the future of decisions that we must take now.

ROBERT N. CHARETTE

CYBER HAS ALWAYS HAD A CLOSE RELATIONSHIP TO SAFETY

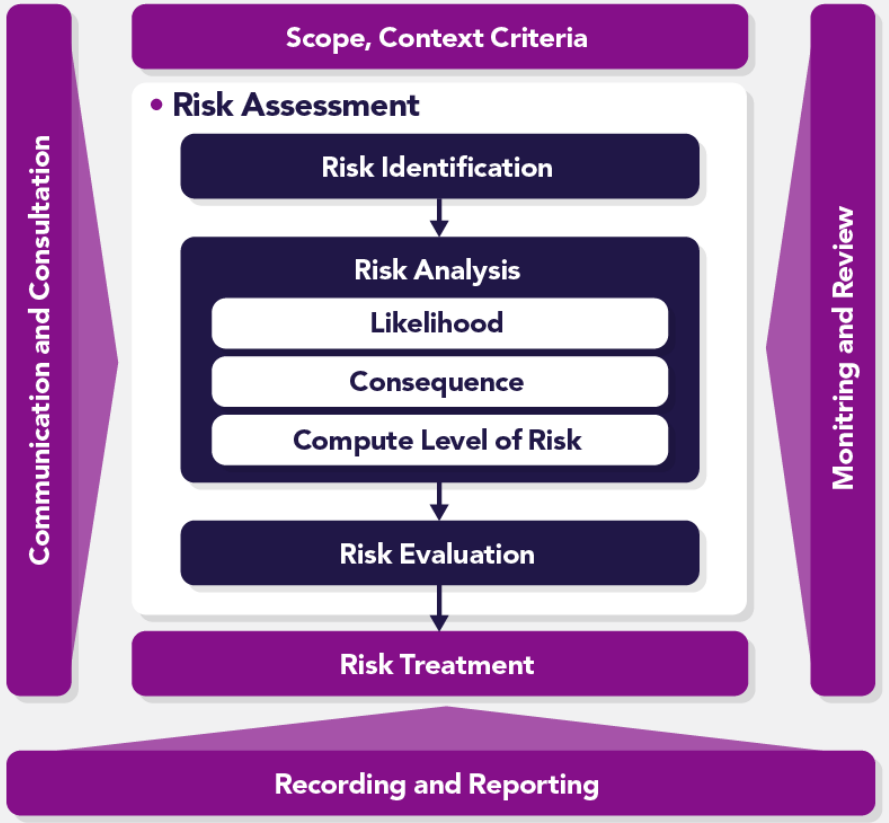
- Both Safety and Cyber are driven by risk based decisions
  - Safety has the concept of hazards with severity of outcome and likelihood of outcome and consequence and mitigations, cyber has the concept of threats, vulnerabilities and controls.
- That risk is always context dependent, and therefore dependent on the mission.
  - Safety hazards that may occur in an urban environment can be different to those in a marine environment.
- Risks are constantly changing
  - COVID was not directly considered a program or safety risk for capabilities until 2020.
  - Lithium batteries in cars is an emerging issue.
  - Cyber risks and threats are ever evolving.
- A decision by a risk owner on what they are tolerant to accepting vs. not. However…
  - ALARP – As low as reasonably practicable.
  - SFAIRP – So far as is reasonably practicable.

CONDUCTING RISK MANAGEMENT

- **Conduct of Risk Management**
  - ISO 31000 Risk Management Standard, provides a consistent approach to the identification, treatment, monitoring and reporting on both cyber and safety risks can be considered.
  - This can then be integrated in both a safety case, or as required, a cyberworthiness assessment.

***Source:**
ISO 31000 Risk Management

00475-7_JS_R1



INTEGRATING CYBER WITH SAFETY

Safety	**ISO31000 Risk Management Standard Process**	**Cyber**
Mission Context	**Scope Context Criteria**	Mission Context
Preliminary Hazard List Hazard Analysis	**Risk Assessment**	Threat Assessment Vulnerability Assessment Risk Assessment
Requirements and Design Verification and Validation Test Program	**Risk Treatment**	Requirements and Design IRAP assessment and Verification and Validation
Safety Case Document	**Recording and Reporting**	Cyberworthiness/Accreditation Documentation

SCOPE, CONTEXT, CRITERIA

Safety: Understanding the “who, what, when, where, why” of the system, its use, and what mission it supports.

Cyber: Understanding the “who, what, when, where, why” of the system, its use, and what mission it supports.

Integrating Cyber within the safety analysis includes:

- Understanding how the system is going to be used and how humans will interact with the system.
- The information flows between different components and the guard rails within the system.
- What mission processes will the system be involved with and support? Can those processes continue if the system is unavailable?

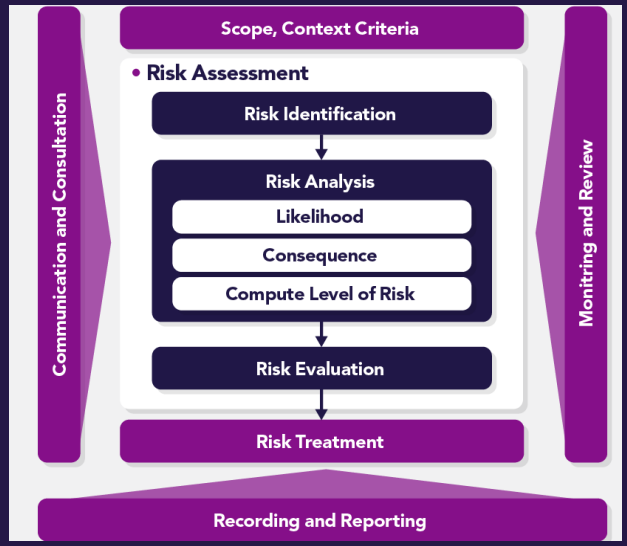
RISK ASSESSMENT

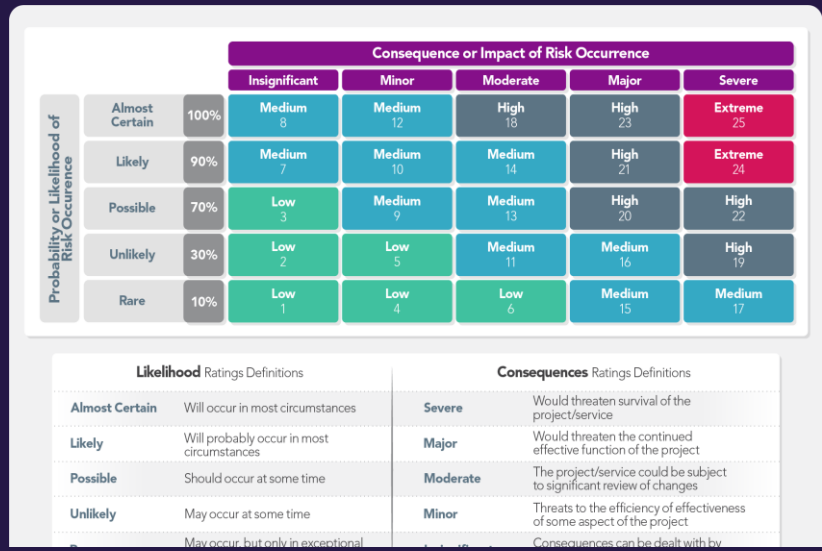
Safety: Conducting a hazard identification and then hazard analysis/risk assessment to give an overall system safety risk perspective.

Cyber: Conducting a cyber threat, vulnerability and risk assessment to give an overall system cyber risk perspective

Integrating Cyber within the safety analysis includes:

- Identifying linkages between software and hardware components that would introduce safety hazards (i.e. the relationship between the autonomous software system and the aircraft's rudder).
- Understanding the guard rails that have been put in place within the autonomous system, and what would occur if those guard rails were removed.





Probability or Likelihood of Risk Occurrence		Consequence or Impact of Risk Occurrence				
		Insignificant	Minor	Moderate	Major	Severe
Almost Certain	100%	Medium 8	Medium 12	High 18	High 23	Extreme 25
Likely	90%	Medium 7	Medium 10	Medium 14	High 21	Extreme 24
Possible	70%	Low 3	Medium 9	Medium 13	High 20	High 22
Unlikely	30%	Low 2	Low 5	Medium 11	Medium 16	High 19
Rare	10%	Low 1	Low 4	Low 6	Medium 15	Medium 17

Likelihood Ratings Definitions		**Consequences** Ratings Definitions	
Almost Certain	Will occur in most circumstances	Severe	Would threaten survival of the project/service
Likely	Will probably occur in most circumstances	Major	Would threaten the continued effective function of the project
Possible	Should occur at some time	Moderate	The project/service could be subject to significant review of changes
Unlikely	May occur at some time	Minor	Threats to the efficiency of effectiveness of some aspect of the project
	May occur, but only in exceptional		Consequences can be dealt with by

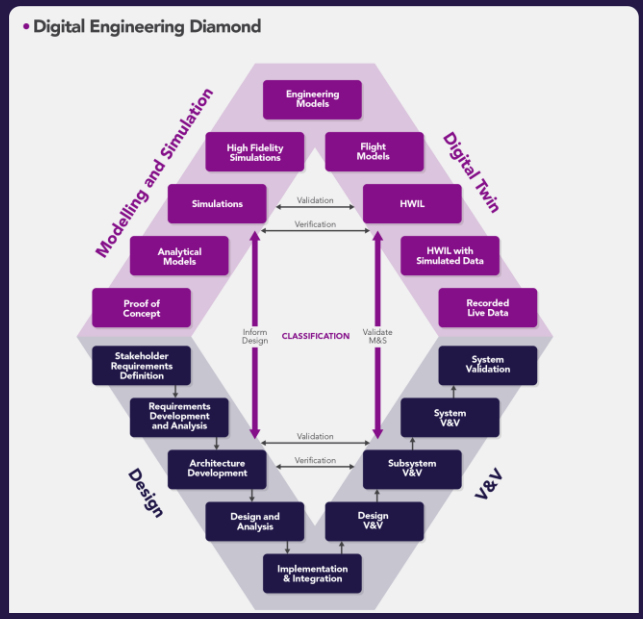
IMPLEMENTATION OF MITIGATIONS

Safety: Identify relevant safety requirements, implement relevant safeguards, and ensure that the safeguards are fit for purpose whilst not impacting the effectiveness of the mission.

Cyber: Identify relevant security requirements, implement relevant controls based on those requirements, and ensure that the controls are fit for purpose whilst not impacting the effectiveness of the mission.

Integrating Cyber within the safety analysis includes:

- Integrating both cyber and safety within traditional systems engineering processes.
- Developers being trained to implement appropriate controls.
- Effective verification of the implementation of controls from a cyber perspective.



IMPLEMENTATION: SUPPLY CHAIN OF AUTONOMOUS SYSTEMS

- Assurance of an autonomous system is to ensure that the system performs its function as the human expects it will
- This also includes verifying that the function can be performed within a safe and secure manner using defined guard rails
- This verification needs to occur at both a software, and a hardware level
- This all requires a robust test and evaluation framework, with built in iterative evaluation cycles.
- Is the autonomous system deterministic or non deterministic?
  - A non deterministic system may be better in operations, but how do you verify?

IMPLEMENTATION: CYBER ARCHITECTURES

- Any cyber attack on the system can modify the systems behaviour and modify the guard rails that are in place.
- An autonomous or operational technology system can be built using traditional cyber architecture principles in mind.
  - Defence in depth
  - Zero Trust
- This also includes any system that is connected to the autonomous system:
  - Base stations
  - Control centres
- The security controls within the autonomous system are equivalent to the antibodies within the human system.

IMPLEMENTATION: VERIFICATION AND VALIDATION

- Verification and Validation activities provide the platform for assurance of the system, its controls and any remaining defects.
- An InfoSec Registered Assessors Program (IRAP) Assessment is essentially a V&V process for cyber controls on a system.
- Testing the platform with scenarios developed from the mission context (Scope, Context and Criteria):
  - In an airspace with a number of objects also in that space.
  - Whilst a given system is disabled or presented with fake information to ensure defence in depth.
- Deterministic systems are “easy” to test. Non deterministic are difficult.

RECORDING AND REPORTING

- **Safety:** A Safety Case consists of a **structured argument**, supported by a body of evidence that provides a compelling, comprehensible and valid case that a system is **safe** for a given application **in a given environment**. – MOD Defence Standard 00-56
- **Cyber:** Cyberworthiness is an **assessment** of the resilience of a system from **cyber attack** within a **given mission context.** – Leidos Australia

Integrating Cyber within the safety analysis includes:

- Developing safety case documentation including cyberworthiness information and assurance baselines.

FEEDBACK LOOP

- In safety, the continuous evaluation of the hazard landscape allows the appropriate controls to be in place at the appropriate time.
- Cyber is no different, you will need to continuously conduct these processes to ensure that you are:
  - secured within a given context.
  - secured against emerging threats.
- The process described above is therefore a continuous cycle – and should be seen as such by both vendors and procurement/acquisition agencies.

Safety doesn’t happen by accident.

AUTHOR UNKNOWN

WHY INTEGRATE CYBER AND SAFETY?

- For autonomous systems, cyber security can equal safety.
  - In order to have a complete safety case, you must understand the cybersecurity risks within a system.
- Safety is a concept more understood by most individuals.
  - Legislative and policy requirements enforce an understanding of safety such as WHS
  - Must be considered in order to meet these requirements.
- Vendors, procurement/acquisition and commanders therefore, should consider how cyber risks have been factored within the safety analysis of their platforms. Assurance of a mission requires that these factors be considered.

SUMMARY

- Having clearly defined and consistent terms can drive unusual linkages (cyber and safety).
- Safety and cyber security can be considered as interdependent within an autonomous systems context.
- Understanding the given mission context for a system enables mission assurance of the platforms.
- Examples of specific mitigations and considerations to be applied within an autonomous context, noting that these need to be considered continuously.

Thank you

IAIN.C.DICKSON@AU.LEIDOS.COM