Space System Survivability

Against AI-Driven Cyberattacks

Applying Structural Security Design Principles for Mission Resilience

Dr. Ron Ross | Dr. Kymie Tan

KAIperShield

NIST SP 800-160 · NASA SunRISE · Mission Resilience

Why This Matters Now

AI-enabled attack tools can now autonomously discover vulnerabilities and generate functional exploits at machine speed.

For space systems — satellites, ground stations, command uplinks — that cannot be rapidly patched once deployed, this changes everything.

Our research on the NASA SunRISE mission demonstrates that structural security engineering — not reactive vulnerability management — is the engineering pivot that preserves mission resilience against determined adversaries.

The Threat Has Changed

NATURE OF EXPLOITATION

Unchanged

- Initial access
- Lateral movement
- Privilege escalation
- Persistence & evasion
- Mission execution

ECONOMICS OF EXPLOITATION

Transformed

- Machine speed, minimal human input
- Autonomous exploit generation
- Full-campaign orchestration at scale
- Compressed time, cost, and expertise
- Accessible to more adversaries

AI changes the speed, scale, and economics of exploitation, not its nature — and that changes everything for defenders.

Mythos-Class Attack Capabilities

AI enables the full offensive campaign — not just exploit generation

Reconnaissance

Autonomously maps entire attack surfaces at scale — identifying exploitable paths across complex software stacks without human direction.

Campaign Planning

Prioritizes high-value targets, sequences multi-stage attack chains, and optimizes intrusion strategies end-to-end.

Adaptive Malware

Modifies behavior in real time in response to defensive measures — evades detection and adjusts tactics dynamically.

Agentic Orchestration

Coordinates multi-stage intrusions across system boundaries without human direction or intervention.

The Space System Attack Surface

1

Ground Segment

Mission operations centers, C2 systems, data processing, and networks.

Primary adversary entry point — compromise grants access to the command uplink.

2

Link Segment

RF and optical paths connecting ground to spacecraft.

Vulnerable to jamming, spoofing, eavesdropping, and unauthorized command injection.

3

Space Segment

Spacecraft bus, payload, onboard computers, and flight software.

On-orbit patching is constrained — vulnerabilities persist far longer than in terrestrial systems.

Mission Threads: End-to-end sequences traversing all three segments create architectural dependencies and control authority relationships. An adversary who maps these gains a blueprint for disrupting mission outcomes — not merely individual components.

Domain Separation

The Architectural Foundation for Space System Defense

“Physical or logical separation of domains with distinctly different protection needs.” — NIST SP 800-160

Susceptibility Reduction

Shields each domain from external influence. Structural isolation prevents adversarial traversal at domain boundaries — independent of whether the attack is detected.

Containment

Prevents a compromised domain from propagating damage outward. A breached IT workstation cannot reach the command uplink — by architecture, not detection.

Domain separation makes adversarial traversal structurally impossible across domain boundaries — not merely detectable after the fact.

Eight Complementary Security Design Principles

NIST SP 800-160 — structural defense at every phase of the attack lifecycle

Least Privilege

Only the access required for each function — limits credential-based escalation

Mediated Access

Policy-based chokepoints at every boundary — violations generate detectable signals

Least Functionality

Eliminate unneeded services — removes entire categories of exploitation

Anomaly Detection

Boundaries define normal — any unauthorized crossing is a high-confidence alert

Defense in Depth

Multiple independent barriers — no single failure causes mission compromise

Least Persistence

Active only during authorized windows — denies persistent adversary footholds

Least Sharing

Minimize shared pathways — eliminates lateral movement routes between domains

Reduced Complexity

Simpler systems have smaller, more analyzable and defensible attack surfaces

The SunRISE Pilot Project

Principles in Practice: A Controlled Experiment

NIST, NASA, and JPL applied SP 800-160 principles to SunRISE — a real NASA mission (6 CubeSats studying solar radio bursts) — using an exact digital twin simulation of the Ground Data System.

Same system · Same attacks · Red/Blue teams separated — did not communicate

Twin A — Traditional Approach

- Vulnerability-centric process
- Compliance-based
- Security controls applied after design
- Security treated as a separate function

Twin B — SSE Approach

- Security design principles from day one
- Security as an engineering property
- Active threat intelligence as design input
- Integrated into architecture — not bolted on

SunRISE: The Results

Twin A (compliance-based) vs. Twin B (NIST SP 800-160)

Data Tampering Detection

Twin A

Weeks

Twin B

Minutes

Mission Data Destruction Detection

Twin A

Weeks

Twin B

Seconds

Malware-Based Incursion

Twin A

Undetected

Twin B

Detected

Security monitoring integrated at design phase: **~2–3 lines of code**
Same capability added post-ATO: **prohibitively expensive**

Implications for Mission-Critical Systems

Beyond space systems — applicable to all 16 CISA critical infrastructure sectors

1 SSE Must Become the Default

Trustworthy secure systems engineering must replace compliance-centered approaches as the standard for all mission-critical systems and high-value assets.

2 Component-Level Security Matters

Security design principles must be applied at the system component level — not only at the system architecture level — to facilitate composite trustworthy secure systems.

3 Evidence-Based Assurance

Trustworthiness is measured by engineering evidence — not compliance attestations. SunRISE demonstrates what that evidence looks like in practice.

4 Laws and Policy Must Evolve

Current acquisition and regulatory frameworks do not consistently require SSE. The SunRISE and SPHEREx experiments are building the empirical case for change.

Conclusion

The Threat:	Mythos-class AI attack tools represent a qualitative escalation — autonomous, machine-speed campaigns change the defender calculus permanently.
The Principle:	Structural security — domain separation and eight complementary SP 800-160 principles — is the only defense that scales against Mythos-class adversaries.
The Evidence:	NASA SunRISE proved it. Detection improved from weeks to seconds. Security integrated at design costs 2–3 lines of code. The same capability post-ATO is prohibitively expensive.

The failure mode is not inadequate technology. It is inadequate engineering.