International Research Journal of Multidisciplinary Sciences ISSN:AWAITED
VOL-1 ISSUE-7 July 2025 PP:32-41

The Impact of Security Threats on Safety-Critical Systems: A Systematic Literature Review and Guidance for Future Research

Abstract

Safety-critical systems (SCS) are integral to modern society, governing operations in domains where failure can result in catastrophic loss of life, significant environmental damage, or substantial economic harm. Historically, the development of these systems has been dominated by rigorous safety engineering methodologies designed to ensure reliability and resilience against random failures. However, the increasing connectivity and sophistication of SCS have exposed them to a new class of threats: intentional, malicious cyber-attacks. This systematic literature review synthesizes existing research on the impact of security threats on SCS. It explores the fundamental conflict between safety and security goals, analyzes the distinct nature of security-induced failures compared to traditional random failures, and examines the consequences across key domains such as automotive, aviation, healthcare, and industrial control systems. The review identifies a critical gap in integrated methodologies that simultaneously address safety and security (cross-domain assurance). Based on the analysis, this paper provides guidance for future research, emphasizing the need for novel risk assessment frameworks (e.g., STPA-Sec), resilient system architectures, runtime monitoring techniques, and the formal verification of combined safety and security properties. The conclusion underscores that the convergence of safety and security is not merely a technical challenge but a fundamental paradigm shift essential for ensuring the trustworthiness of next-generation critical infrastructure.

Keywords: Safety-Critical Systems, Cybersecurity, Systematic Literature Review, Risk Assessment, Resilient Architecture, Cross-Domain Assurance, STPA, Cyber-Physical Systems

1. Introduction

Safety-critical systems (SCS) are defined as systems whose failure or malfunction could result in one or more of the following outcomes: death or serious injury to people, loss or severe damage to equipment/property, or environmental harm (Leveson, 2011). For decades, the engineering of SCS has been a mature discipline, grounded in standards like ISO 26262 (automotive), DO-178C (avionics), and IEC 61508 (industrial). These frameworks are exceptionally effective at mitigating risks from random hardware failures and systematic software errors through processes like Failure Mode and Effects Analysis (FMEA), Fault Tree Analysis (FTA), and the establishment of rigorous Safety Integrity Levels (SILs).

The 21st century has witnessed the rapid transformation of once-isolated SCS into highly connected, software-intensive Cyber-Physical Systems (CPS). While this connectivity enables unprecedented functionality, efficiency, and data exchange (e.g., autonomous vehicle-to-everything (V2X) communication, smart grid management, remote patient monitoring), it also creates a vast and vulnerable attack surface. Malicious actors are no longer a theoretical concern; incidents like the Stuxnet worm, which targeted Iranian nuclear centrifuges (Langner, 2011), the demonstrated remote

hijacking of a Jeep Cherokee (Miller & Valasek, 2015), and ransomware attacks on hospitals have proven the materialization of these threats.

This convergence creates a fundamental conflict: traditional safety engineering assumes failures are random or accidental, while security threats are intelligent, adaptive, and deliberate. A security breach can intentionally trigger a failure mode that safety analyses deemed sufficiently improbable to ignore. This paradigm shift means that safety cannot be assured without considering security.

1.1 Problem Statement and Research Objectives

The core problem is that existing engineering processes for SCS are ill-equipped to handle the unique challenges posed by malicious threats. The impact of security threats extends beyond confidentiality and integrity breaches to directly compromise the primary safety goal of SCS: avoiding harm. This paper aims to:

1. Systematically review literature on the impact of security threats on SCS.
2. Analyze the interplay and conflicts between safety and security engineering goals.
3. Identify the unique characteristics of security-induced failures.
4. Synthesize findings across major SCS domains.
5. Provide a guidance framework for future research to bridge the safety-security gap.

1.2 Methodology

This paper employs a systematic literature review (SLR) methodology, following the guidelines established by Kitchenham and Charters (2007). The process involved:

Planning: Defining research questions and a review protocol.

Conducting: Identifying primary research databases (IEEE Xplore, ACM Digital Library, Scopus, Web of Science) and selecting studies using keywords: ["safety-critical systems" AND cybersecurity], ["security" AND "safety" AND "impact"], ["cyber-physical systems" AND security threat], ["resilience" AND "safety-critical"].

Screening: Filtering results based on relevance, publication date (primarily 2010-2023), and peer-review status.

Synthesizing: Extracting and thematically analyzing data from the selected studies to identify key trends, challenges, and proposed solutions.

2. The Safety-Security Dichotomy and Convergence

To understand the impact of security threats, one must first appreciate the inherent tensions between the disciplines of safety and security engineering.

2.1 Fundamental Goals and Conflicts

Safety is concerned with the absence of catastrophic consequences on the user(s) and the environment. It is primarily focused on preventing accidental harm caused by system failures. Its core strategy is management of uncertainty and probabilistic risk assessment (Ericson, 2005).

Security is concerned with the protection of system assets from unauthorized access, modification, or destruction. It is focused on preventing intentional harm caused by a malicious adversary. Its core strategy is deterrence, prevention, detection, and response to threats (Anderson, 2008).

The conflict arises in their implementation (Young & Leveson, 2014):

Connectivity vs. Isolation: Security often advocates for isolation and minimal interfaces to reduce attack surfaces. Safety, particularly in modern systems, often requires connectivity for diagnostics, updates, and coordinated control (e.g., vehicle platooning).

Timeliness vs. Overhead: Safety functions must execute within strict deadlines. Cryptographic security controls can introduce latency and computational overhead, potentially violating these timing constraints.

Availability: Safety requires high availability ("fail-operational" in many cases). Security actions, such as shutting down a system upon detecting an intrusion, directly conflict with this requirement, potentially creating a denial-of-service condition.

Risk Assessment: Safety uses probabilistic models (e.g., failure rates). Security must reason about the capabilities, resources, and intentions of an intelligent adversary, which is inherently non-probabilistic.

2.2 The Need for Convergence

Despite these conflicts, safety and security are two sides of the same coin: dependability . A system cannot be truly safe if it is vulnerable to attacks that can cause it to fail dangerously. Conversely, a security measure that causes a safety-critical function to fail is unacceptable. This necessitates an integrated approach often termed "cross-domain assurance" or "safety-security co-engineering."

3. The Unique Nature of Security-Induced Failures in SCS

Security threats impact SCS in ways that differ fundamentally from traditional random failures, exacerbating their severity (Krotofil & Cárdenas, 2013).

Intentionality and Adaptability: An adversary learns, adapts, and specifically targets the system's weakest points. They can exploit multiple, seemingly insignificant vulnerabilities in concert to achieve a cascading effect that would be astronomically improbable by chance alone.

Common Cause Failures: Safety engineering relies on redundancy (e.g., triple modular redundancy) to mask random failures. A skilled attacker can identify and compromise all redundant channels simultaneously, defeating this primary safety mechanism.

Stealth and Persistence: Attacks can be designed to remain undetected for long periods, manipulating system data to hide their presence. This allows the attacker to cause gradual damage or lie in wait for a critical moment to strike, making diagnosis and recovery immensely difficult.

Targeting the "Smart" Components: Attacks focus on the software and communication layers—the system's "nervous system"—rather than its physical hardware. This allows for sophisticated attacks like sensor spoofing, command injection, and logic bombs that are difficult to anticipate with traditional FMEAs.

4. Domain-Specific Impacts of Security Threats

The impact of security threats manifests differently across various SCS domains. This section provides a synthesized overview based on the reviewed literature.

4.1 Automotive Systems

Modern vehicles are "computers on wheels," with up to 100 ECUs connected via internal networks (CAN, LIN, FlexRay). The impact of breaches is dire.

Threats: Remote exploitation via infotainment, telematics, or Bluetooth; sensor spoofing (e.g., fooling cameras/LiDAR); malicious firmware updates.

Impact: Miller and Valasek (2015) famously demonstrated remote control over steering, braking, and acceleration. Security threats directly translate to life-threatening safety failures: loss of vehicle control, collision, and fatalities.

Research Focus: Intrusion Detection Systems (IDS) for CAN buses, secure gateway architectures, and over-the-air (OTA) update security.

4.2 Aviation (Avionics)

Aircraft are highly complex SCS with a long history of safety. Their increasing connectivity for passenger Wi-Fi, air traffic control links, and maintenance data loads creates new risks.

Threats: Compromise of ground systems, spoofing of GPS or ADS-B signals, potential attacks on flight control systems via infected maintenance laptops.

Impact: While core flight systems are often isolated, attacks on navigation or communication systems could lead to mid-air collisions, controlled flight into terrain, or guidance to incorrect locations. The 2015 report by the U.S. GAO highlighted cybersecurity as an emerging threat to flight safety.

Research Focus: Airworthiness security standards (e.g., DO-326A/ED-202A), partitioning to ensure core avionics isolation, and verifying security of connected components.

4.3 Medical Devices

Healthcare SCS, such as implantable cardiac devices (ICDs), insulin pumps, and patient monitors, are increasingly wireless and networked.

Threats: Remote hijacking of devices to deliver lethal shocks (e.g., ICDs) or incorrect drug doses (e.g., insulin pumps); ransomware locking clinicians out of critical patient data during surgery.

Impact: Direct physical harm or death to patients. Halperin et al. (2008) demonstrated the ability to wirelessly compromise an ICD. Attacks also threaten patient privacy and can disrupt hospital operations.

Research Focus: Secure communication protocols for medical devices, anomaly detection in device behavior, and pre-market cybersecurity guidance from regulators like the FDA.

4.4 Industrial Control Systems (ICS) and Critical Infrastructure

This domain includes SCADA systems managing power grids, water treatment plants, oil and gas pipelines, and manufacturing facilities.

Threats: The Stuxnet attack is the quintessential example, causing physical damage to centrifuges by manipulating programmable logic controllers (PLCs) (Langner, 2011). Other threats include ransomware (e.g., Colonial Pipeline attack) and reconnaissance attacks mapping operational technology (OT) networks.

Impact: Widespread power outages, failure of safety instrumented systems (SIS), environmental catastrophes (e.g., dam overflow, chemical leak), and massive economic disruption.

Research Focus: Network segmentation between IT and OT, deep packet inspection for industrial protocols, and resilience strategies that allow for continued safe operation under attack.

5. Guidance for Future Research: Bridging the Gap

The literature review reveals that while the problem is well-identified, integrated solutions are still in their infancy. Future research must move beyond siloed approaches. Here is a guidance framework for key research directions:

5.1 Integrated Safety and Security Risk Assessment

Traditional FTA and FMEA are inadequate for modeling intelligent threats. Future work must adopt and extend holistic techniques that can model both random and malicious causes.

STPA-Sec: Systems-Theoretic Process Analysis for Security is an extension of STPA that treats security as a control problem and can identify insecure control actions and loss scenarios resulting from malicious actions (Young & Leveson, 2014). Research is needed to create automated tools and formalized models for this methodology.

Combined FTA-FTBA: Integrating Fault Tree Analysis with Fault Tree-Based Attack (FTBA) trees can provide a unified view of how a top-level failure event (e.g., "loss of braking") can be triggered by either a random fault or a successful attack path.

Quantitative Framework: Developing a unified metric that can express risk incorporating both safety (probability of failure) and security (level of effort for an attacker) parameters.

5.2 Resilient System Architecture and Design

Systems must be designed from the ground up to maintain safety even in the presence of a security breach (i.e., be "resilient").

Adaptive Security: Research into security controls that can dynamically adjust their level of protection (and thus their overhead) based on the current operational mode of the SCS. For example, during a critical maneuver, cryptographic verification might be simplified to meet timing constraints, while being strengthened during non-critical phases.

Cyber-Physical Resilience: Architectures that can detect compromises and gracefully degrade functionality to a "safe state" rather than failing catastrophically. This involves research into redundant, diverse, and heterogeneous components that are difficult for an attacker to simultaneously compromise.

Runtime Monitoring and Intrusion Detection: Moving beyond network-based IDS to model-based intrusion detection that checks the physical consistency of the CPS (e.g., does the commanded actuator movement make sense given the current vehicle dynamics and sensor readings?).

5.3 Formal Methods and Verification

The high assurance required for SCS demands rigorous verification.

Co-Verification of Properties: Extending formal methods tools to verify both safety and security properties simultaneously. For instance, using model checking to verify that a security policy does not violate a timing constraint essential for safety.

Verifying Machine Learning Components: As AI/ML becomes prevalent in autonomous SCS (e.g., perception systems), research is urgently needed to verify their robustness against adversarial examples—specially crafted inputs designed to cause the ML model to make a catastrophic error.

5.4 Human Factors and Organizational Aspects

Training: Developing cross-disciplinary training for engineers who are literate in both safety and security principles.

Supply Chain Security: Research into methods for vetting the security of third-party components and software libraries that are integrated into SCS, a major source of vulnerability.

Regulatory Evolution: Research to inform policymakers and standards bodies (ISO, IEC) on how to update existing safety standards (e.g., ISO 26262) to formally incorporate security considerations, perhaps leading to a new generation of unified standards.

6. Conclusion

The increasing connectivity and complexity of safety-critical systems have fundamentally altered their risk landscape. This systematic literature review confirms that intentional, malicious security threats pose a severe and distinct danger to systems designed to protect human life and critical infrastructure. The impact is no longer theoretical; it is demonstrated across automotive, aviation, healthcare, and industrial domains, where a security breach can be a direct cause of a safety failure.

The historical separation between safety and security engineering is a dangerous anachronism. Safety, which concerns itself with accidental failures, cannot be assured in a world where systems are exposed to intelligent adversaries who can deliberately induce those very failures. The core challenge lies in reconciling the conflicting goals and methodologies of these two disciplines.

This paper has outlined the unique characteristics of security-induced failures and synthesized the domain-specific impacts. Most importantly, it provides a guidance framework for future research, highlighting four critical pathways: 1) the development of integrated risk assessment techniques like STPA-Sec, 2) the design of inherently resilient system architectures, 3) the advancement of formal co-verification methods, and 4) addressing human and organizational factors.

The path forward requires a concerted, cross-disciplinary effort. Researchers, engineers, and policymakers must move beyond their traditional silos to develop a new paradigm of "cross-domain assurance." The convergence of safety and security is not merely a technical inconvenience; it is the foundational requirement for building trustworthy systems in the 21st century. The cost of inaction is measured not in data breaches, but in lives lost.

7. References

1. Anderson, R. (2008). Security Engineering: A Guide to Building Dependable Distributed Systems (2nd ed.). Wiley.

2. Ericson, C. A. (2005). Hazard Analysis Techniques for System Safety . John Wiley & Sons.

3. Government Accountability Office (GAO). (2015). AVIATION CYBERSECURITY: FAA Should Fully Implement Key Practices to Strengthen Its Oversight of Avionics Risks . GAO-15-741.

4. Halperin, D., Heydt-Benjamin, T. S., Ransford, B., Clark, S. S., Defend, B., Morgan, W., ... & Fu, K. (2008). Pacemakers and Implantable Cardiac Defibrillators: Software Radio Attacks and Zero-Power Defenses. In IEEE Symposium on Security and Privacy (pp. 129-142).

5. Kitchenham, B., & Charters, S. (2007). Guidelines for performing Systematic Literature Reviews in Software Engineering . EBSE Technical Report, Keele University and Durham University.

6. Krotofil, M., & Cárdenas, A. A. (2013). Resilience of Process Control Systems to Cyber-Physical Attacks. In Nordic Conference on Secure IT Systems (pp. 166-182). Springer.

7. Langner, R. (2011). Stuxnet: Dissecting a Cyberwarfare Weapon. IEEE Security & Privacy , 9(3), 49–51.

8. Leveson, N. G. (2011). Engineering a Safer World: Systems Thinking Applied to Safety . The MIT Press.

9. Miller, C., & Valasek, C. (2015). Remote Exploitation of an Unaltered Passenger Vehicle. Black Hat USA , 2015.

10. Young, W., & Leveson, N. G. (2014). An integrated approach to safety and security based on systems theory. Communications of the ACM , 57(2), 31-35.

11. Relevant Standards: ISO 26262, DO-178C, IEC 61508, DO-326A/ED-202A, ISO/SAE 21434.