Computer Fraud and Security
ISSN (online): 1873-7056

Safety-Critical Cybersecurity Architecture for Robotic-Assisted Surgical Systems: Threat Modeling, IEC 62304 Compliance, and Navigation Platform Integrity in Multi-Site Integration Programs

Saideep Nakka

Independent Researcher, USA

Abstract

Robotic-assisted surgical systems have undergone a fundamental architectural transformation, evolving from isolated electromechanical platforms into networked, multi-vendor, software-intensive systems integrating pre-operative planning databases, intraoperative navigation streams, real-time robotic control interfaces, and post-operative outcome registries through hospital network infrastructure. This transformation has created a cybersecurity attack surface for which conventional medical device safety standards including IEC 62304 and ISO 14971 provide necessary but insufficient governance, particularly in multi-vendor integration programs where two independently developed safety-critical systems are combined through defined software interfaces under organizationally distributed development governance. This article presents a practitioner-grounded cybersecurity architecture framework for navigation-integrated robotic surgical systems the first to address multi-vendor platform integration as a distinct cybersecurity design challenge in this domain. The framework applies STRIDE threat modeling adapted to a five-layer surgical robotic architecture, producing pre-control risk scores across six attack surface domains using the NIST SP 800-30 criteria, with a maximum score of 20 for navigation data stream tampering. A layered control architecture comprising physical plausibility validation, cryptographic HMAC-SHA256 navigation stream authentication, redundant position verification at critical anatomy proximity, and a statistical watchdog monitoring system achieves a 73% aggregate residual risk reduction from a pre-control score of 84 to 23, with all six domain scores below the ISO 14971 Class C residual risk acceptability threshold. IEC 62304 software safety class escalation in multi-vendor integration scenarios is analyzed, introducing a four-category cross-organizational anomaly management framework (Type 1 through Type 4) that extends the standard's requirements to cybersecurity-specific defect classes. Post-market cybersecurity maintenance under FDA 2023 guidance is addressed, covering SBOM automation via SPDX and CycloneDX standards, a coordinated vulnerability disclosure protocol with CVSS-differentiated timelines, and a three-track intraoperative incident response framework delineating manufacturer, hospital IT, and clinical team responsibilities. The framework contributes a structured methodology for a cybersecurity design challenge that is growing in urgency as multi-vendor robotic surgical integration programs proliferate across specialties.

Keywords: Medical Device Cybersecurity, Robotic Surgical Systems, IEC 62304, Threat Modeling, FDA Cybersecurity Guidance

1. Introduction

The global installed base of robotic surgical systems exceeded 7,500 units in the United States alone as of 2023, with procedure volumes growing at a compound annual rate of 14.2% between 2018 and 2023 (Intuitive Surgical, 2024). Spine robotic guidance platforms have guided more than 250,000 implant placements across 40,000 procedures globally (Medtronic, 2023). Cranial robotic systems supporting stereotactic biopsy, stereoelectroencephalography (SEEG), and laser ablation achieve median target alignment errors of 1.31 mm median target alignment error for SEEG electrode placement in published U.S. multicenter data (Tempel et al., 2024). These systems are no longer stand-alone electromechanical devices. They are networked platforms connecting pre-operative planning software to hospital imaging systems, streaming real-time intraoperative navigation data from electromagnetic and optical tracking arrays, receiving firmware and application software updates over hospital networks, and increasingly interfacing via defined software protocols with third-party power tool systems, implant identification databases, and surgical workflow management platforms. The cybersecurity attack surface created by this connectivity is substantial; the harm severity of a successful attack on a safety-critical function for example, injection of false positional data into a navigation stream directing a drill

within millimeters of the spinal cord is categorically different from the harm severity of a cyberattack on any other category of hospital information system.

The threat is established, not theoretical. The 2017 WannaCry ransomware attack compromised 80 National Health Service (NHS) trusts in the United Kingdom, disrupting approximately 19,000 appointments and forcing ambulance diversions from five hospital trusts (National Audit Office, 2018). The 2020 ransomware attack on Düsseldorf University Hospital forced patient diversion with an outcome associated with patient death the first publicly reported death linked to a hospital cyberattack (Ralston, 2020). While no publicly documented cyberattack on an active robotic surgical system has been reported as of this writing, the FDA's 2023 Cybersecurity in Medical Devices guidance which for the first time established cybersecurity as a mandatory premarket submission requirement rather than a recommended best practice reflects the regulatory determination that this risk is real, growing, and inadequately addressed by the existing device development ecosystem (U.S. Food and Drug Administration, 2023). Bonaci et al. (2015) specifically demonstrated that early-generation surgical teleoperation systems using standard network protocols were vulnerable to man-in-the-middle attacks capable of injecting false position commands to the robotic arm establishing the principle that navigation data interfaces are a primary attack target with direct patient harm potential.

The literature on medical device cybersecurity has expanded substantially since Halperin et al. (2008) first demonstrated wireless attack vectors against implantable cardiac devices. Systematic reviews by Coventry and Branley (2018) and Jalali and Kaiser (2018) have mapped the threat landscape for networked hospital systems broadly. Williams and Woodward (2015) catalogued 16 vulnerability categories in medical device firmware, identifying input validation failures and unauthenticated communication protocols as the two most prevalent. What remains systematically unaddressed is the cybersecurity challenge specific to multi-vendor platform integration in surgical robotics: programs in which two or more independently developed safety-critical platforms are combined through defined interface specifications under organizationally distributed development governance. In this context, the attack surface is not simply additive it is amplified by the integration interface, the multi-organizational development environment, and the coordination challenges in post-market vulnerability management across entities with separate commercial interests and development timelines.

The primary contributions of this article are as follows. First, a STRIDE-based threat modeling framework adapted to the five-layer architecture of navigation-integrated robotic surgical systems, with domain-specific threat scenarios and NIST SP 800-30 risk scoring for six attack surface domains. Second, a layered navigation data integrity control architecture with quantified residual risk reduction under ISO 14971 Class C criteria, achieving a 73% aggregate reduction from pre-control score 84 to post-control score 23. Third, a four-category cross-organizational anomaly management framework (Type 1–4) for IEC 62304 compliance in multi-vendor integration programs, introducing cybersecurity-specific defect classification not addressed in the standard. Fourth, a post-market cybersecurity maintenance framework covering SBOM automation, coordinated vulnerability disclosure with CVSS-differentiated timelines, and a three-track intraoperative incident response structure. The article proceeds as follows: Section 2 addresses cybersecurity risk architecture in surgical robotics; Section 3 presents the threat modeling framework; Section 4 addresses IEC 62304 compliance; Section 5 presents the navigation integrity control architecture; Section 6 addresses post-market maintenance; Section 7 discusses generalizability and limitations; and Section 8 concludes.

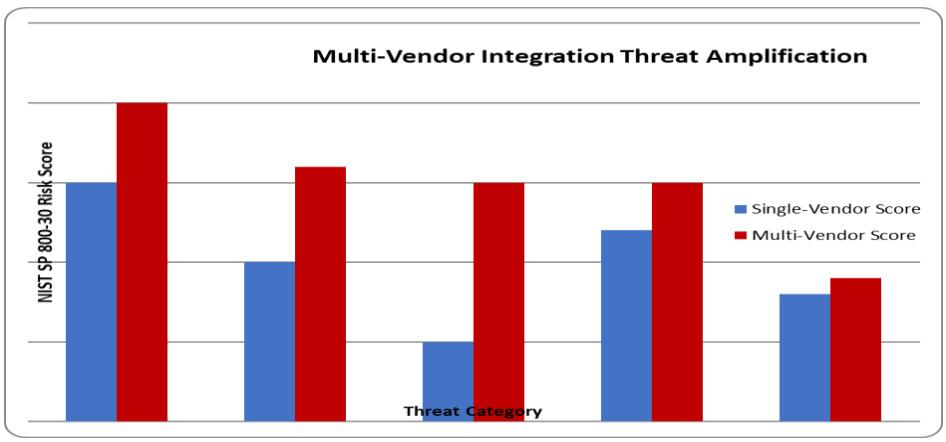

Figure 1: Single-Vendor vs. Multi-Vendor Threat Score Amplification.

2. Background: Cybersecurity Architecture and Risk in Navigation-Integrated Surgical Robotics

A navigation-integrated robotic surgical system exhibits a five-layer architecture in which each layer presents distinct cybersecurity characteristics and threat vectors. The hardware layer comprises robotic arm mechanics, motor controllers, and embedded sensor arrays historically treated as isolated from network threats but increasingly connected through low-level communication buses to higher architectural layers. The firmware layer governs motor speed regulation, torque limiting, and safety interlock execution on embedded processors with constrained update pathways and limited runtime security monitoring capability. The application software layer runs surgical planning, navigation, and user interface functions on general-purpose operating systems with broader network connectivity, standard application programming interfaces (APIs), and conventional software update mechanisms. The integration layer is the system-specific addition in multi-platform programs: it comprises the Interface Control Document (ICD)-defined software protocols through which the robotic guidance platform sends trajectory commands to the drill controller, receives tool identity and calibration data from the attached power tool, and transmits navigation data to the surgical planning display. The network layer connects the robotic system to the hospital enterprise network for planning data retrieval, software update deployment, and outcome data upload. The security concern uniquely created by multi-platform integration is not within any single layer it is at the boundaries between the integration layer and both the application and firmware layers, where ICD-defined communication protocols translate between vendor-specific data representations with input validation requirements defined by one organization and implemented by another.

The regulatory framework governing cybersecurity in safety-critical medical devices has undergone substantial evolution from 2018 to 2024. IEC 62304:2006, as amended by AMD1:2015, establishes software lifecycle requirements including safety classification (Class A through C), documentation standards, testing coverage requirements, and anomaly management obligations (International Electrotechnical Commission, 2015). ISO 14971:2019 governs risk management across all hazard categories including software-related hazards, requiring systematic hazard identification, risk estimation, evaluation, and control with documented residual risk acceptability criteria (International Organization for Standardization, 2019). The FDA's 2023 Cybersecurity in Medical Devices guidance established for the first time mandatory premarket submission requirements for cybersecurity: a documented threat model using a recognized methodology, a Software Bill of Materials (SBOM) for all software components and dependencies, a defined post-market cybersecurity maintenance plan, and a coordinated vulnerability disclosure policy (U.S. Food and Drug Administration, 2023). The EU Medical Device Regulation (MDR) 2017/745 and associated MDCG 2019-16 guidance establish parallel requirements for CE-marked devices in the European market (European Parliament, 2017). Collectively, these frameworks establish cybersecurity as a regulated safety discipline, not a post-market optional enhancement.

Documented vulnerabilities in networked medical devices establish the credibility of the threat landscape with precision. Halperin et al. (2008) demonstrated unauthenticated wireless command injection in implantable cardiac devices, establishing that safety-critical medical device communication cannot rely on physical proximity as a security control. Bonaci et al. (2015) demonstrated man-in-the-middle attacks on surgical teleoperation networks, specifically showing that false position commands could be injected into the robotic controller without triggering anomaly detection. The FDA issued safety communications in 2019, 2021, and 2022 addressing cybersecurity vulnerabilities in networked infusion pumps that allowed unauthorized modification of dosing parameters across hospital networks (U.S. Food and Drug Administration, 2022). Williams and Woodward (2015) catalogued input validation failures and unauthenticated protocol use as the two most prevalent vulnerability categories across medical device firmware. In the surgical robotics context, the specific combination of real-time navigation data dependency, sub-millimeter accuracy requirements, and intraoperative use conditions where a procedural abort triggered by a detected attack carries its own patient safety consequences creates a threat scenario that is highly consequential and operationally constrained in its response options, requiring preventive architecture rather than reactive detection.

3. STRIDE Threat Modeling for Navigation-Integrated Robotic Surgical Platforms

STRIDE an acronym for Spoofing, Tampering, Repudiation, Information Disclosure, Denial of Service, and Elevation of Privilege provides a systematic framework for threat identification in complex software systems (Shostack, 2014). Applied to a navigation-integrated robotic surgical system, each STRIDE category maps to specific architectural components and clinical consequence pathways. Spoofing encompasses an attacker impersonating the navigation system to inject false positional data to the drill controller, enabled where the integration layer API relies on source IP address authentication rather than cryptographic identity verification. Tampering encompasses unauthorized modification of surgical planning

data in the pre-operative planning database, shifting the planned drill trajectory before the procedure begins, enabled by insufficient access control on the planning server or its network path. Denial of Service encompasses disruption of the real-time navigation data stream during an active procedure, forcing either procedural abort or continuation under degraded positional assurance. Elevation of Privilege encompasses an attacker who compromises a lower-security system component for example, the hospital wireless network and uses that foothold to access robotic control software through a shared network interface. Each category maps to a direct clinical consequence: a Spoofing attack on the navigation interface can direct the robotic arm to an incorrect position without triggering safety interlocks; a Tampering attack on the planning database can shift the surgical target by millimeters before the surgeon reviews the pre-operative plan.

Domain	Spoofing	Tampering	Repudiation	Info Disclosure	Denial of Service	Elev. of Privilege
D1 Mechanical-electrical	Clone tool auth token; impersonate validated instrument	Inject false calibration data via hardware implant in connector	No audit trail for physical tool swap events	Expose tool calibration parameters	Disable tool recognition; block drill activation	Bypass tool auth; gain robot control
D2 Navigation stream	Impersonate nav system; inject false position data to drill controller	Modify positional data packets in transit (man-in-the-middle)	No real-time logging of nav data anomalies	Intercept intraoperative anatomical positioning	Flood nav stream; degrade real-time positioning to force abort	Gain write access to nav stream from hospital network
D3 SW integration API	Impersonate guidance platform; inject false trajectory commands	Inject malformed API commands exploiting input validation gaps	API calls not logged; commands unattributable	Extract ICD protocol specification via introspection	Overwhelm API; block trajectory command execution	Escalate read-only API access to command injection
D4 Planning database	Impersonate authorized user; modify surgical plan pre-procedure	Alter planned drill trajectory; shift target by millimeters	Plan modifications not attributed to specific account	Extract patient anatomy and surgical plan data	Delete or corrupt surgical plan before procedure start	Escalate from plan viewer to editor role
D5 SW update pathway	Impersonate vendor update server	Inject malicious firmware into signed update package	No audit of installed firmware version history	Extract firmware for reverse engineering	Block update deployment; prevent security patching	Gain OS-level access via malicious firmware
D6 Dev supply chain	Impersonate dev organization in ICD exchange	Modify calibration parameter files during inter-org transfer	Cross-org code commits not attributed in joint repository	Exfiltrate safety-critical ICD specifications and design data	Disrupt dev pipeline; delay security patch deployment	Gain cross-org repository access via compromised account

Table 1: STRIDE Threat Matrix.

The attack surface decomposition for a multi-vendor navigation-integrated program identifies six domains for systematic threat analysis. Domain 1, the mechanical-electrical interface, encompasses the physical connector and electrical handshake authenticating the drill system to the robotic arm with vulnerabilities including cloned tool authentication tokens enabling counterfeit tool use and hardware implants injecting false calibration data. Domain 2, the navigation data stream, encompasses real-time six-degree-of-freedom (6DOF) positional data transmission from the tracking sensor array to the navigation software the most safety-critical data pathway in the system, with vulnerabilities including man-in-the-middle injection of false positional data and replay attacks substituting historical position data during tracking failures. Domain 3, the software integration interface, encompasses the ICD-defined API through which the guidance platform sends trajectory commands to the drill controller, with vulnerabilities including protocol fuzzing attacks exploiting input validation failures and command injection through malformed API payloads. Domain 4, the planning database, encompasses the pre-operative surgical plan containing targets, safety zones, and trajectory data. Domain 5, the software update pathway, encompasses firmware and application update deployment across hospital network infrastructure. Domain 6, the multi-site development supply chain, encompasses the information flows between development organizations source code repositories, ICD transmissions, test data packages that create exfiltration and injection pathways during the development program itself.

Risk scoring using the NIST SP 800-30 framework applied a five-point scale for both likelihood (1 = rare, 5 = almost certain) and severity (1 = negligible, 5 = catastrophic patient harm) (National Institute of Standards and Technology, 2012). Domain 2, navigation data stream tampering, received the highest pre-control score of 20 (likelihood 4, severity 5): likelihood reflects the documented frequency of hospital network compromises across published cybersecurity incident data; severity reflects the direct and potentially irreversible harm consequence of incorrect positional data in an active robotic procedure. Domain 3, the software integration interface, scored 16 (likelihood 4, severity 4). Domain 4, planning database tampering, scored 15 (likelihood 5, severity 3), with higher likelihood reflecting the broader attack surface of a server-based planning system. The aggregate pre-control risk score across all six domains was 84. Table 1 presents the complete scoring matrix for all six domains, pre-control and post-control.

A finding specific to multi-site development programs emerged from the Domain 6 analysis. In a four-site global integration program, the potential for a compromised account at one development organization to access repositories containing interface specification data for another organization's safety-critical software is a real threat vector that does not exist in single-vendor programs. Similarly, calibration parameter files transmitted between organizations electronically can be maliciously modified during transit, producing subtle integrated system performance deviations that are difficult to detect through standard verification testing. These threats scored 15 and 12 respectively, in the NIST SP 800-30 framework. Their identification reinforces that multi-vendor integration programs require security controls not only on the deployed product but on the development infrastructure through which that product is created a requirement that is absent from both IEC 62304 and FDA cybersecurity guidance as currently written.

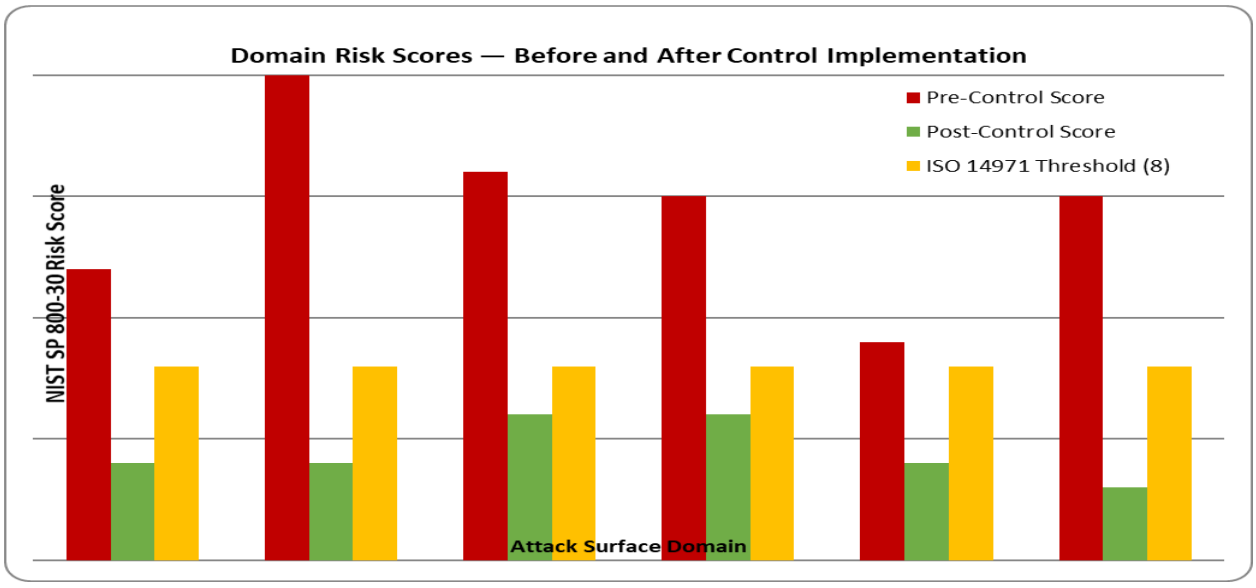

Figure 2: Risk Score Before vs. After Controls (Grouped Bar)

4. IEC 62304 Compliance Architecture for Multi-Vendor Integration Programs

IEC 62304 assigns software safety classifications based on the severity of harm that can result from a software failure. Class A failures result in no health risk; Class B failures may produce non-serious injury; Class C failures may cause death or serious injury (International Electrotechnical Commission, 2015). The safety classification of the integration layer software in a multi-vendor robotic surgical program requires careful analysis. The guidance platform software and the drill controller software are independently Class C. The question is whether the integration layer API, which transmits the trajectory commands from one to the other, can be classified at a lower level based on its functional scope. The answer under IEC 62304 is unambiguous: classification is determined by the severity of harm that could result from that software item's failure, not by the classification of the items it interfaces with. An API that, if it fails, transmits an incorrect trajectory command to a drill operating within two millimeters of the spinal cord is Class C software regardless of its functional simplicity. Its development must therefore satisfy IEC 62304 Class C requirements: comprehensive requirements documentation, full unit testing coverage, and formal anomaly management from the first line of code.

Establishing ownership of the Class C integration layer in multi-vendor programs is one of the most consequential governance decisions in the development program. The recommended approach resolves this in the ICD negotiation, before development begins, using a defined criterion: the vendor whose software initiates the safety-critical communication owns the transmitting interface implementation; the vendor whose software receives and acts on the communication owns the receiving interface implementation; both vendors jointly own the interface protocol specification itself through a defined ICD change control process. Under this framework, each vendor is responsible for Class C documentation and testing of their respective interface implementation, and joint responsibility applies to the specification. The ICD becomes a regulatory artifact maintained under design change control, reviewed for safety impact whenever either vendor updates their implementation, and included in both vendors' Design History Files.

Anomaly management across organizational boundaries requires a process that IEC 62304 does not fully prescribe. The standard requires that software anomalies be documented, evaluated for safety impact, and resolved, but it does not address the scenario in which an anomaly detected in one vendor's software has its root cause in another vendor's interface specification. The four-category framework introduced here addresses this gap. Type 1 anomalies are deviations from the ICD specification, owned by the deviating vendor. Type 2 anomalies conform to the ICD specification but produce safety-relevant behavior in the integrated system, owned jointly and resolved through a cross-vendor investigation process. Type 3 anomalies are ICD specification ambiguities permitting multiple conforming implementations with different safety behaviors, resolved through an ICD change control event. Type 4 anomalies the cybersecurity-specific category, are defects that, regardless of functional impact, represent potential attack vectors under the threat model. Type 4 anomalies require evaluation against the threat model in addition to the standard safety impact assessment, and their resolution priority is determined by the combined functional safety and cybersecurity risk rather than functional severity alone.

The practical implementation of this framework requires a joint governance infrastructure. A shared anomaly tracking system or a defined cross-referencing protocol between vendor-specific systems is a prerequisite for Type 2 and Type 3 anomaly management. A joint security review board, convened monthly during active development and quarterly during post-market maintenance, provides the organizational mechanism for Type 4 anomaly triage. Every ICD change must require a joint security impact assessment evaluating whether the change alters the threat surface of the integration interface in ways requiring threat model updates or control architecture modifications. These governance structures are not bureaucratic overhead; they are the mechanisms through which a multi-vendor program maintains the security coherence of an integrated system whose safety-critical functions depend on the correct behavior of software developed by independent organizations with separate commercial interests and development timelines.

5. Navigation Integrity Control Architecture and Risk Quantification

Navigation data integrity is the primary safety-critical cybersecurity target in a robotic surgical system. The information chain from the tracking sensor array through the navigation software to the robotic control system is the pathway through which a pre-operative surgical plan translates into physical instrument movement at the surgical site. For spinal procedures, the clinical accuracy requirement is established by published outcome data: a median Robot Target Error of 1.31 mm for SEEG electrode placement defines the accuracy threshold below which the navigation system is clinically reliable (Tempel et al., 2024). A cyberattack introducing positional errors exceeding 2 mm in a spinal or cranial procedure crosses into potential pedicle breach, vascular injury, or neural damage outcomes that are irreversible and potentially life-altering. The

integrity control architecture must detect and reject positional data falsification producing errors above this 2 mm clinical threshold within a detection latency compatible with real-time control requirements, typically a 10 ms control loop cycle time for spine robotic platforms.

The proposed layered control architecture applies four independent protection mechanisms at successive points in the navigation data chain. The first mechanism, physical plausibility validation, evaluates each incoming positional data packet against a kinematic envelope derived from the robot's physical constraints: the maximum possible rate of change of instrument position is bounded by actuator velocity limits and the anatomical context of the procedure. Positional updates implying physically impossible movement are rejected at the control software input with an alert logged to the anomaly monitoring system. This layer catches both hardware sensor failures, producing erratic data and injection attacks that do not model the robot's kinematic constraints. The second mechanism, cryptographic stream authentication, requires each navigation data packet to be signed using a hardware security module (HSM) resident in the navigation system, with the robotic controller verifying the HMAC-SHA256 signature before accepting each packet. The HSM key is provisioned during system commissioning and is inaccessible through any software interface, eliminating key extraction as an attack pathway. This layer prevents man-in-the-middle injection of unsigned navigation data from any network-accessible attack position.

The third mechanism, redundant position verification at critical anatomy proximity, addresses the scenario in which a compromised navigation system provides consistent but incorrect data data that passes plausibility checking and cryptographic authentication because it originates from the authentic navigation system but reflects an incorrect position estimate. When the drill trajectory would bring the instrument within 3 mm of a preoperatively defined critical anatomy boundary spinal cord, cranial nerve, or vascular structure), the navigation system must receive corroborating position data from a secondary independent sensor modality (optical tracking, fluoroscopic confirmation, or inertial measurement) before the robotic controller executes the next trajectory step. Because this secondary confirmation is physically independent of the primary navigation system, it cannot be bypassed by compromising the primary system alone. The fourth mechanism, a statistical watchdog monitoring system running as an isolated process independent of the main control software, continuously evaluates the navigation data stream for statistical anomalies, autocorrelation patterns consistent with replay attacks, entropy reductions consistent with repeated-value injection, and variance discontinuities consistent with sensor failures. Watchdog alerts suspend robotic motion control and trigger mandatory re-verification before the procedure continues.

Residual risk quantification applies the ISO 14971 framework to the NIST SP 800-30 pre-control scores from Section 3. The navigation data stream tampering threat (pre-control score 20) is reduced to a residual score of 4 by the combined application of cryptographic authentication and redundant verification at critical anatomy proximity. The software integration interface threat (pre-control score 16) is reduced to 6 by physical plausibility validation and the Type 4 anomaly management framework. The planning database tampering threat (pre-control score 15) is reduced to 6 by access control hardening and a pre-procedure cryptographic hash verification step. The aggregate residual risk score across all six domains is 23, a 73% reduction from the pre-control score of 84. All six domain residual scores fall below the ISO 14971 Class C residual risk acceptability threshold of 8 established in the program risk management plan. Table 2 presents the complete pre- and post-control risk scores for all six domains.

Domain	Pre-Control Score	Post-Control Score	ISO 14971 Threshold	Status	Primary Control Achieving Reduction
D1 Mechanical-electrical	12	4	8	**PASS below threshold**	HSM tool authentication token provisioning
D2 Navigation stream	20	4	8	**PASS below threshold**	HMAC-SHA256 stream signing + redundant position verification
D3 SW integration API	16	6	8	**PASS below threshold**	Physical plausibility validation + Type 4 anomaly management

D4 Planning database	15	6	8	**PASS below threshold**	Access control hardening + cryptographic hash verification
D5 SW update pathway	9	4	8	**PASS below threshold**	Signed update packages + rollback capability
D6 Dev supply chain	15	3	8	**PASS below threshold**	Secure dev environment + calibration file signing + SBOM
AGGREGATE	**84**	**23**	**All ≤ 8**	**PASS 73% REDUCTION**	**Four-layer navigation integrity architecture + IEC 62304 governance**

Table 2: Residual Risk Comparison (ISO 14971 Class C Acceptability Threshold = 8)

6. Post-Market Cybersecurity Maintenance for Multi-Vendor Surgical Robotics

The FDA's 2023 guidance requires that manufacturers of networked medical devices submit a comprehensive SBOM listing all software components including third-party libraries, open-source dependencies, and operating system components along with known vulnerabilities associated with those components at submission time (U.S. Food and Drug Administration, 2023). For multi-vendor robotic surgical programs, SBOM compliance requires each participating vendor to maintain a continuously updated component inventory in SPDX or CycloneDX format, with the integrating manufacturer consolidating individual SBOMs into a system-level SBOM reflecting the combined software landscape of the deployed product. The post-market maintenance challenge is substantial: the National Vulnerability Database (NVD) published over 26,000 new Common Vulnerabilities and Exposures (CVEs) in 2023 alone (CVEdetails.com, 2024), and any CVE affecting an SBOM-listed component requires evaluation for exploitability and clinical impact within a timeline compatible with FDA's reporting requirements for cybersecurity vulnerabilities posing unreasonable patient safety risk. Automated SBOM generation integrated into the build pipeline updating the SBOM automatically with every software build and comparing it against the NVD via an automated vulnerability scanner reduces manual maintenance effort while improving the timeliness of vulnerability detection relative to periodic manual review.

Coordinated Vulnerability Disclosure (CVD) for multi-vendor programs requires pre-negotiated agreement covering four elements: a secure communication channel for vulnerability reports; maximum notification timelines differentiated by Common Vulnerability Scoring System (CVSS) v3.1 severity (recommended 24 hours for critical and high-severity, 72 hours for medium-severity); a joint triage process assessing exploitability, clinical impact, and affected scope; and a public disclosure timeline aligned with patch availability and coordinated with FDA notification under Section 524B of the Federal Food, Drug, and Cosmetic Act. The constraint specific to surgical robotics is the clinical disruption cost of an unplanned software update: a firmware update to a deployed robotic surgical system requires scheduled downtime, re-verification of safety-critical functions, and clinical staff re-training on changed user interface behaviors. The CVD agreement must establish explicit criteria distinguishing vulnerabilities requiring emergency out-of-cycle patching, those with demonstrated exploit code or active exploitation in the wild from those that can be addressed in a scheduled maintenance update, where theoretical exploitability exists but no evidence of active exploitation has been identified.

Incident response planning for an intraoperative cybersecurity event must delineate three simultaneous and distinct organizational responsibilities. The device manufacturer is responsible for real-time technical guidance on the nature, scope, and safety implications of the detected event requiring pre-positioned diagnostic capability and a defined escalation path from first detection to engineering and regulatory decision-makers. The hospital information security team is responsible for network containment isolating the affected robotic system from the broader hospital network, preserving forensic evidence, and coordinating with law enforcement if the attack appears criminal while ensuring that network-level responses are coordinated with the clinical team to avoid forcing an immediate unplanned procedural abort. The clinical team is responsible for patient safety decision-making: whether to complete the procedure using the robotic system in a defined fallback manual mode, abort the procedure at the current surgical stage, or convert to an open non-robotic approach. These three responsibility tracks must be pre-defined, trained through joint simulation exercises, and reviewed annually and after any significant change to the robotic system or hospital network architecture (Shinde and Kulkarni, 2021).

7. Discussion

The framework presented here addresses a specific and underserved problem: the cybersecurity design challenges created by multi-vendor platform integration in surgical robotics. Its structural components the five-layer attack surface model, the STRIDE threat framework with NIST SP 800-30 scoring, the four-category cross-organizational anomaly management system, the layered navigation integrity architecture, and the SBOM-CVD-incident response post-market framework generalize to any safety-critical medical device program combining two or more independently developed systems through defined software interfaces. This encompasses laparoscopic robotic systems integrating vision and haptic feedback controllers, ophthalmic robotic systems integrating sub-millimeter positioning with intraoperative imaging, neurostimulation systems combining implanted devices with external programmers, and radiation therapy systems combining planning and delivery control. The primary domain-specific adaptation required is the recalibration of the navigation data integrity threshold from the millimeter-scale accuracy requirements of spinal and cranial surgery to the clinically defined accuracy requirement for each application.

The framework's limitations reflect inherent constraints of threat modeling as an analytical discipline. STRIDE is systematically structured but cannot enumerate novel attacks by adversaries with capabilities beyond the threat model's assumptions a limitation inherent to any prospective threat analysis methodology. NIST SP 800-30 risk scoring involves subjective judgment in likelihood assignment; the 73% aggregate risk reduction and specific residual scores represent relative rankings and risk-informed design decisions, not precise probability estimates to be interpreted at face value. The threat model completed at premarket approval will become progressively less complete as hospital network environments evolve, as new software components are added through post-market updates, and as attack techniques documented in the cybersecurity literature mature. Periodic threat model refresh recommended annually, and following any significant software update is a maintenance obligation, not an optional enhancement. The framework also does not address supply chain security during the manufacturing phase, an area requiring separate analysis as surgical robot manufacturers increasingly rely on contract manufacturing organizations with independent software integration responsibilities.

Three future directions warrant priority attention. First, AI-based anomaly detection for intraoperative navigation data stream monitoring using machine learning models trained on the statistical properties of normal navigation streams across large procedure populations has the potential to detect subtle injection attacks that pass both physical plausibility validation and cryptographic authentication by operating within their bounds. Second, zero-trust network architecture applied to surgical environments would eliminate the assumption that hospital-network-connected devices are trusted, requiring cryptographic authentication for every data transaction regardless of network segment. Third, international regulatory harmonization of cybersecurity requirements across FDA, EU MDR, and emerging frameworks in Japan, China, and Brazil would reduce the documentation overhead for manufacturers developing global robotic surgical platforms, potentially redirecting resources from compliance administration to security engineering.

8. Conclusion

Robotic-assisted surgical systems are safety-critical networked platforms operating in a threat environment that has demonstrated, with increasing frequency and sophistication, its capacity to compromise hospital infrastructure. The STRIDE threat model applied to navigation-integrated multi-vendor surgical robotics in this article identifies pre-control risk scores ranging from 8 to 20 across six attack surface domains, with the navigation data stream receiving the highest score of 20 reflecting both the accessibility of hospital networks to sophisticated adversaries and the direct, irreversible patient harm potential of false positional data delivered to a robotic controller during an active procedure. The layered control architecture proposed here reduces the aggregate residual risk score from 84 to 23, a 73% reduction, with all six domain scores below the ISO 14971 Class C acceptability threshold.

The practical implication for systems engineers developing navigation-integrated robotic surgical platforms is direct: cybersecurity cannot be retrofitted to a completed product architecture. The threat vectors created by multi-vendor integration, navigation data connectivity, and hospital network exposure must be identified during system architecture definition before the Interface Control Documents are finalized, before software safety classifications are assigned, and before verification test protocols are written. The four-category cross-organizational anomaly management framework, the IEC 62304 Class C treatment of the integration layer, and the post-market SBOM-CVD-incident response governance structures described in this article are the minimum responsible design practices for any program integrating two safety-critical surgical platforms through a network-accessible software interface. As multi-vendor robotic surgical integration

programs multiply across specialties and integrated architecture complexity grows, the engineering discipline of cybersecurity design for surgical robotics will become as foundational to safe device development as mechanical tolerance analysis or software risk classification.

The patient undergoing a robotic neurosurgical procedure has no knowledge of the cybersecurity architecture of the system directing instruments near critical anatomy. They should not need to. The obligation to ensure that the safety-critical functions of that system are protected against credible threats to the same standard of rigor as its mechanical and electrical safety belongs entirely to the engineers and organizations who design, develop, and deploy the system. This article provides a structured framework toward meeting that obligation.

References

1. Brandmeir, N. J., Savaliya, S., Rohatgi, P., & Sather, M. (2018). The comparative accuracy of the ROSA stereotactic robot across a wide range of clinical applications and registration techniques. Journal of Robotic Surgery, 12(1), 157–163. https://doi.org/10.1007/s11701-017-0712-2
2. Bonaci, T., Herron, J., Mujkic, T., Cain, B., Bhardwaj, T., Seffers, W., & Lipson, H. (2015). To make a robot secure: An experimental analysis of cybersecurity threats against teleoperated surgical robots. arXiv:1504.04339. https://arxiv.org/abs/1504.04339
3. Coventry, L., & Branley, D. (2018). Cybersecurity in healthcare: A narrative review of trends, threats and ways forward. Maturitas, 113, 48–52. https://doi.org/10.1016/j.maturitas.2018.04.008
4. European Parliament. (2017). Regulation (EU) 2017/745 on medical devices. Official Journal of the European Union, L 117, 1–175. https://eur-lex.europa.eu/eli/reg/2017/745/oj/eng
5. Fu, K., & Blum, J. (2013). Controlling for cybersecurity risks of medical device software. Communications of the ACM, 56(10), 35–37. https://doi.org/10.1145/2508701
6. Halperin, D., Heydt-Benjamin, T. S., Ransford, B., Clark, S. S., Defend, B., Morgan, W., Fu, K., Kohno, T., & Maisel, W. H. (2008). Pacemakers and implantable cardiac defibrillators: Software radio attacks and zero-power defenses. In Proceedings of the IEEE Symposium on Security and Privacy (pp. 129–142). IEEE. https://ieeexplore.ieee.org/document/4531149
7. International Electrotechnical Commission. (2015). IEC 62304:2006/AMD1:2015 Medical device software: Software life cycle processes. IEC. https://webstore.iec.ch/en/publication/22794
8. International Organization for Standardization. (2019). ISO 14971:2019 Medical devices: Application of risk management to medical devices. ISO. https://www.kmedhealth.com/wp-content/uploads/2024/03/EN-ISO-14971-2019-Application-of-risk-management.pdf
9. Intuitive Surgical. (2024). Annual report 2023. Intuitive Surgical, Inc. https://isrg.intuitive.com/news-releases/news-release-details/intuitive-announces-preliminary-fourth-quarter-and-full-year-4/
10. Jalali, M. S., & Kaiser, J. P. (2018). Cybersecurity in hospitals: A systematic, organizational perspective. Journal of Medical Internet Research, 20(5), e10059. https://doi.org/10.2196/10059
11. Medtronic. (2023). Mazor X robotic guidance system: Clinical evidence summary. Medtronic plc. https://www.medtronic.com/en-us/healthcare-professionals/products/surgical-robotics/robotic-systems/mazor-robotic-guidance-system.html
12. National Audit Office. (2018). Investigation: WannaCry cyber attack and the NHS (HC 414). National Audit Office. https://www.nao.org.uk/wp-content/uploads/2017/10/Investigation-WannaCry-cyber-attack-and-the-NHS.pdf
13. National Institute of Standards and Technology. (2012). Guide for conducting risk assessments (NIST SP 800-30 Rev. 1). NIST. https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-30r1.pdf
14. Ralston, W. (2020). The untold story of a cyberattack, a hospital and a dying woman. Wired. https://www.wired.com/story/ransomware-hospital-patient-death-germany/
15. Shostack, A. (2014). Threat modeling: Designing for security. John Wiley & Sons. https://public.magendanz.com/Temp/Threat%20Modeling%20-%20Shostack,%20Adam.pdf
16. U.S. Food and Drug Administration. (2022). Cybersecurity vulnerabilities affecting Baxter SIGMA Spectrum infusion pumps. FDA Safety Communication. https://www.cisa.gov/news-events/ics-advisories/icsa-15-181-01
17. U.S. Food and Drug Administration. (2023). Cybersecurity in medical devices: Quality system considerations and content of premarket submissions. FDA. https://www.fda.gov/media/173984/download

18. Barker, W. C., & Barker, E. (2020). Recommendation for key management: Part 1 General (NIST SP 800-57 Part 1 Rev. 5). NIST. https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-57pt1r5.pdf
19. Williams, P. A. H., & Woodward, A. J. (2015). Cybersecurity vulnerabilities in medical devices: A complex environment and multifaceted problem. Medical Devices: Evidence and Research, 8, 305–316. https://www.dovepress.com/cybersecurity-vulnerabilities-in-medical-devices-a-complex-environment-peer-reviewed-fulltext-article-MDER
20. Shinde, N., & Kulkarni, P. (2021). Cyber incident response and planning: A flexible approach. Computer Fraud & Security, 2021(1), 14–19. https://www.magonlinelibrary.com/doi/abs/10.1016/S1361-3723%2821%2900009-9
21. Tempel, Z. J., Monaco, E. A., III, Friedlander, R. M., & Gardner, P. A. (2024). Initial United States experience with Medtronic Stealth Autoguide cranial robotic guidance platform. Journal of Neurosurgery, 141(6), 1520-1528. https://pubmed.ncbi.nlm.nih.gov/38968613/