Umsetzungshinweise zum Baustein: MRIS.bd.1 Mythos-resistente Härtung des ISMS

Version 1.2 | Stand: 14.08.2026 | Autor: Richard Peddi

1 Einleitung

Diese Umsetzungshinweise konkretisieren die Anforderungen des Bausteins MRIS.bd.1 Mythos-resistente Härtung des ISMS. Grundlage sind das Framework „Mythos-resistente Informationssicherheit" [MRIS] und der zugehörige Implementation Guide [MRISIG]. „Mythos" bezeichnet dabei die Klasse agentischer Frontier-KI-Modelle mit autonomer Schwachstellen-Entdeckung und Exploit-Kettenbau, unabhängig vom Anbieter.

Die Maßnahmen folgen einem gemeinsamen Grundprinzip: Schutzwirkung soll aus harten strukturellen Barrieren stammen (kryptografische Unmöglichkeit, Unveränderlichkeit, nachweisbare Trennung, begrenzte Rechte), nicht aus Angreiferreibung, die unter KI-Bedingungen nicht mehr trägt. Jede Maßnahme benennt daher neben der Umsetzung auch die Messung der Wirksamkeit und typische Fehler. Die in den Maßnahmen genannten Produkte sind unverbindliche Beispiele und keine Empfehlung.

2 Maßnahmen

Im Folgenden sind spezifische Maßnahmen für die Anforderungen des Bausteins MRIS.bd.1 Mythos-resistente Härtung des ISMS aufgeführt.

Alle Maßnahmen (gekennzeichnet mit M) sind aufsteigend nummeriert und korrespondieren mit den entsprechenden Anforderungen (gekennzeichnet mit A).

2.1 Maßnahmen zum Baustein MRIS.bd.1 Mythos-resistente Härtung des ISMS

MRIS.bd.1.M1 Neubewertung der Wirksamkeitsannahmen bestehender Sicherheitsmaßnahmen (B)

Grundlage der Neubewertung ist das MRIS-Vier-Kategorien-Raster mit der Leitfrage, ob die Schutzwirkung einer Maßnahme aus einer harten strukturellen Barriere stammt oder nur aus Angreiferreibung: standfest (29 Controls der ISO/IEC 27002:2022), teilweise degradiert (37), reine Reibung (4) und nicht betroffen (23). Die MRIS-Bewertungsmatrix (Anhang A, zusätzlich als maschinenlesbare YAML-Fassung verfügbar) kann als Startpunkt übernommen und an den eigenen Kontext angepasst werden; die Klassifikation ist kontextabhängig und keine allgemeingültige Festlegung [MRIS].

Das Vorgehen priorisiert die als reine Reibung und als teilweise degradiert eingestuften Maßnahmen für eine Sofort-Neubewertung. Die Ergebnisse fließen in die Risikoanalyse ein — insbesondere die Verschiebung der Eintrittswahrscheinlichkeit durch den Wegfall der Kapazitätsbarriere und die kollabierten Zeitparameter (siehe Baustein, Kap. 2) — sowie in das Statement of Applicability. Je degradierter Maßnahme wird dokumentiert entschieden: Härtung (über die Maßnahmen M2 bis M14), Kompensation oder begründete Risikoübernahme. Wichtig ist die Lesart „selektiv degradiert", nicht „obsolet": Klassische Maßnahmen verlieren in einzelnen Wirksamkeits-Dimensionen, behalten in anderen ihre volle Schutzwirkung — die korrekte Reaktion ist Härtung, nicht Ersatz.

Die Neubewertung wird anlassbezogen wiederholt, insbesondere bei wesentlichen Änderungen der Bedrohungslage (neue Modell-Fähigkeitsklassen, Offenlegungswellen, veränderte Angriffsmuster). Messung: Anteil der neu bewerteten Annex-A-Maßnahmen (Zielwert 100 %) und dokumentierte Entscheidungen je degradierter Maßnahme. Typischer Fehler: Die Neubewertung wird als Einmalprojekt behandelt — ohne definierte Wiederholungsauslöser veraltet sie mit der nächsten Fähigkeitsstufe der Modelle.

MRIS.bd.1.M2 Einsatz phishing-resistenter Multi-Faktor-Authentisierung (B)

Organisatorisch wird die Authentisierungsrichtlinie so gefasst, dass phishing-resistente Verfahren nach FIDO2/WebAuthn für privilegierte, extern erreichbare und besonders schützenswerte Zugänge verbindlich sind und je Zugangsklasse die zugelassenen Verfahren festgelegt werden. Für bestehende Legacy-MFA-Verfahren (SMS, einfache Push-Bestätigung) wird ein terminierter Ablösungsplan erstellt.

Technisch werden Passkeys oder Hardware-Token als Mindestform ausgerollt. Die Wirksamkeit beruht auf der Origin-Bindung von WebAuthn: Ein auf einer Phishing-Seite erbeuteter Authentisierungsnachweis ist für den echten Dienst wertlos, unabhängig von der Qualität KI-generierter Phishing-Inhalte. Besondere Aufmerksamkeit gilt den Registrierungs- und Wiederherstellungsprozessen: Diese Randprozesse sind der bevorzugte Umgehungspfad und werden gegen Social Engineering abgesichert (Identitätsprüfung, Vier-Augen-Prinzip bei privilegierten Konten). Zu beachten ist ferner, dass eine Kompromittierung des Endgeräts nach erfolgreicher Anmeldung durch das Verfahren nicht verhindert wird; Session-Schutz und Endpoint-Härtung bleiben eigenständige Aufgaben.

Messung: Anteil der Anmeldungen über phishing-resistente Verfahren, getrennt nach Zugangsklassen; Zielwert ist die vollständige Abdeckung der privilegierten und extern erreichbaren Zugänge. Typische Fehler: Das Verfahren wird eingeführt, aber SMS bleibt als Fallback aktiv; die Wiederherstellung neuer Authentisierungsmittel läuft über einen ungesicherten Helpdesk-Prozess.

MRIS.bd.1.M3 Vorhalten unveränderlicher Datensicherungen mit regelmäßigen Wiederherstellungstests (B)

Die Datensicherung wird nach dem 3-2-1-1-0-Modell aufgebaut: drei Kopien auf zwei Medientypen, mindestens eine Kopie offsite und mindestens eine unveränderlich (Immutability/Object-Lock) oder offline, null Fehler beim periodischen Wiederherstellungstest. Die Backup-Infrastruktur erhält eigene, von der Produktionsumgebung getrennte IAM-Pfade, damit kompromittierte Produktions-Administratorkonten die Sicherungen nicht löschen oder verändern können.

Wiederherstellungstests werden mindestens vierteljährlich durchgeführt und dokumentiert; geprüft wird die tatsächliche Rückspielbarkeit definierter geschäftskritischer Datenbestände, nicht nur die Existenz der Sicherung. Unveränderliche Sicherungen sind eine der wirksamsten Hartbarrieren gegen Ransomware und agentische Datenintegritätsangriffe, weil ihre Schutzwirkung nicht von der Erkennungsgeschwindigkeit abhängt.

Messung: Erfolgsquote der vierteljährlichen Wiederherstellungstests (Zielwert 100 %); Nachweis über Testprotokolle, Aufbewahrungs- und Sperreinstellungen sowie die getrennte Rechtevergabe. Typische Fehler: Die Unveränderlichkeit ist konfigurierbar und mit denselben Administratorrechten wieder abschaltbar; Wiederherstellungstests prüfen nur Einzeldateien statt vollständiger Systeme.

MRIS.bd.1.M4 Aufbau einer verhaltensbasierten Angriffserkennung mit Ereigniskorrelation (B)

Die Erkennung wird von signaturorientierten Einzel-Alarmen auf verhaltensbasierte Muster und Kill-Chain-Korrelation über mehrere Ereignisse und Zeitfenster umgestellt. Grundlage sind MITRE-ATT&CK-basierte Erkennungsregeln mit messbarer Abdeckung; als Richtwert gilt eine Abdeckung

von mindestens 60 % der zehn wichtigsten Techniken der eigenen Branche, gemessen etwa mit DeTT&CT und bestätigt durch Angriffssimulationen (z. B. Atomic Red Team) [ATTACK]. Erkennungs-Schwerpunkte gegen KI-typische Vorgehensweisen sind Living-off-the-Land-Aktivitäten (PowerShell-Encoding, ungewöhnliche Prozessketten aus Office-Anwendungen, auffällige curl-/wget-Aufrufe) und getarnte Command-and-Control-Kanäle (DNS-Tunneling, HTTPS mit langen Sleep-Intervallen, missbrauchte Cloud-APIs).

Threat-Hunting wird als eigenständige Funktion mit dokumentierter Methodik etabliert (z. B. PEAK oder TaHiTI); als Richtwert gelten 0,5 FTE bis 500 Mitarbeitende und mindestens zwölf hypothesenbasierte Hunts pro Jahr mit ATT&CK-Zuordnung. Für Institutionen ohne eigenes SOC sind Managed-Detection-and-Response-Dienste eine valide Alternative.

Messung: gemessene und durch Tests bestätigte ATT&CK-Abdeckung sowie Zahl und Ergebnis der Hunts. Zu beachten: Detection ist nicht Prevention. Low-and-slow-Angriffe oberhalb der Korrelations-Zeitfenster und gut getarnte Agenten innerhalb legitimer Verhaltens-Baselines bleiben Restrisiko; die Maßnahme wirkt nur in Kombination mit den strukturellen Barrieren (M2, M3, M5). Typischer Fehler: Es werden Werkzeuge beschafft, aber keine Abdeckung gemessen — die Erkennungslücken bleiben unbekannt.

MRIS.bd.1.M5 Einführung von Workload-Identitäten und identitätsbasierter Zugriffskontrolle (B)

Die Umstellung erfolgt in drei Stufen statt als Big-Bang-Migration. Stufe 1 (privilegierte Workloads): kryptografisch verifizierbare, kurzlebige Identitäten (z. B. SPIFFE/SPIRE) für Service-Konten mit erweiterten Rechten sowie mTLS auf den drei bis fünf kritischsten Service-zu-Service-Pfaden. Stufe 2 (produktiver Mainstream): Workload-Identität für alle produktiven Microservices und Zero Trust Network Access (ZTNA) für privilegierten Remote-Zugriff. Stufe 3 (Flächendeckung): Ausdehnung auf Entwicklungs- und Testumgebungen sowie identitätsbasierte Mikrosegmentierung.

KI-Agenten sind ein Sonderfall: Coding-Agenten und agentische Workflows erhalten eigene Workload-Identitäten mit zeitlich begrenzten, auf die jeweilige Fähigkeit zugeschnittenen Berechtigungen pro Werkzeugaufruf — keine persönlichen Entwickler-Zugangsdaten (siehe M7). Langlebige, im Klartext gespeicherte gemeinsame Geheimnisse werden auf kritischen Pfaden eliminiert; sie sind der klassische Treibstoff für Lateral Movement.

Messung: Anteil der produktiven Dienst-zu-Dienst-Verbindungen mit Workload-Identität und mTLS (Richtwerte: Stufe 1 mindestens 80 % der privilegierten Service-Workloads, Stufe 2 mTLS-Pflicht für mindestens 95 % des Ost-West-Verkehrs) sowie die Zeit bis zum Entzug einer Identität. Typische Fehler: mTLS wird eingeführt, aber die Autorisierung erfolgt weiterhin über IP-Adressen; die Identitäten sind langlebig und werden nie rotiert.

MRIS.bd.1.M6 Erstellung eines kryptografischen Inventars und einer Post-Quantum-Migrationsstrategie (B)

Zunächst wird ein zentrales Inventar aller eingesetzten Algorithmen, Schlüssellängen, Protokolle, Zertifikate und Bibliotheken erstellt und nach Schutzbedarf und Vertraulichkeitsdauer der Daten priorisiert. Auf dieser Basis wird eine Migrationsstrategie zu quantensicheren Verfahren dokumentiert, mit definierten Auslösern, Ressourcen, Übergangspfaden und Erfolgskriterien. Der Bedrohungsbezug ist das Muster „heute abgreifen, später entschlüsseln": Der Datenabfluss von heute bestimmt das Entschlüsselungsrisiko von morgen, weshalb Daten mit langer Vertraulichkeitsdauer zuerst migriert werden.

Technisch erfolgt die Umstellung auf die finalisierten NIST-Standards FIPS 203 (ML-KEM) für den Schlüsselaustausch sowie FIPS 204 (ML-DSA) und FIPS 205 (SLH-DSA) für Signaturen [FIPS];

Verfahrens- und Schlüssellängen-Empfehlungen enthält BSI TR-02102 [TR02102], das Migrationsvorgehen NIST SP 1800-38 [SP180038]. Im Übergang werden hybride Verfahren eingesetzt, die klassische und quantensichere Kryptografie kombinieren. Crypto-Agility wird strukturell verankert: Verschlüsselung wird so gekapselt, dass Verfahren ohne tiefe Eingriffe in die Anwendungen austauschbar sind. Institutionen ohne eigene Entwicklung steuern das Thema über Inventarüberblick und Beschaffung, indem sie den Migrationsfahrplan ihrer Anbieter abfragen und quantensichere Verfahren vertraglich einfordern.

Messung: Anteil der erfassten und nach Migrationspriorität eingestuften Verfahren; Toleranzgrenze: keine langlebig vertraulichen Daten ohne Migrationsplan. Typische Fehler: Eine Strategie ohne vollständiges Inventar; flächendeckende Hybrid-Einführung ohne Priorisierung nach Vertraulichkeitsdauer; fest im Code verankerte Verschlüsselung, die jeden Verfahrenswechsel zum Anwendungsprojekt macht.

MRIS.bd.1.M7 Geregelter Einsatz von KI-Agenten (B)

Die Governance produktiver KI-Agenten umfasst fünf Dimensionen:

- Harness-Audit: Prompts, Werkzeug-Definitionen, Retrieval-Pipelines und Eskalationslogik werden mit derselben Code-Review-Disziplin behandelt wie Produktivcode — Versionierung im Source-Repository, signierte Releases, Tests gegen Prompt-Injection und Confused-Deputy-Angriffe vor der Produktivsetzung.
- Blast-Radius-Begrenzung: Berechtigungen pro Werkzeugaufruf mit explizit definiertem Umfang, Rate-Limits pro Agent-Identität, Circuit-Breaker bei ungewöhnlichen Aktionsfolgen und verbindliche menschliche Freigabeschwellen für irreversible Aktionen (Produktions-Deployment, Datenlöschung, Ausgaben oberhalb definierter Grenzen).
- Menschliche Kontrolle: Jede Agent-Sitzung hat eine benannte verantwortliche Person mit Abschaltmöglichkeit; alle Aktionen werden mit Korrelation zur auslösenden menschlichen Identität protokolliert, Aufbewahrung mindestens zwölf Monate.
- Lieferketten-Inventar agentischer Komponenten: MCP-Server (Quelle, Update-Kanal), IDE-Erweiterungen mit KI-Funktion (Freigabeliste, kontrollierte Updates), Skills und Rules-Dateien (Versionierung, Code-Review) sowie externe Agent-Frameworks.
- Prüfung vor Produktivsetzung: dokumentierte Scope-Definition, Bedrohungsmodell für den Anwendungsfall, Rollback-Plan und Monitoring-Plan mit definierten Anomalie-Schwellwerten.

In der ersten Phase werden Inventarisierung, Protokollierung und Rechtebegrenzung priorisiert — diese drei Elemente adressieren bereits den größten Teil des Risikos; vollständige Harness-Penetrationstests folgen abhängig von der Werkzeug-Verfügbarkeit. Messung: Anteil der produktionsnahen Agenten mit dokumentiertem Bedrohungsmodell und festgelegtem Schadensradius (Zielwert 100 %), Protokollierungs-Abdeckung für Aktionen mit Schreib- oder Netzwerkrechten (Zielwert 100 %), Zeit bis zum Entzug kompromittierter Agent-Identitäten (Zielwert unter fünf Minuten). Typischer Fehler: Agenten laufen unter persönlichen Benutzerkonten der Entwickelnden — jede Aktion ist dann weder begrenzbar noch sauber zuordenbar.

MRIS.bd.1.M8 Erzeugung von Software-Stücklisten und Build-Provenance-Nachweisen (S)

Die Stücklisten-Erzeugung wird automatisch in die Build-/CI-Pipeline integriert, in einem etablierten maschinenlesbaren Format — CycloneDX (ECMA-424) oder SPDX (ISO/IEC 5962) —, einschließlich der transitiven Abhängigkeiten. Je Softwareversion entsteht eine eigene, versioniert aufbewahrte Stückliste. Die Komponenten werden automatisiert gegen Schwachstellendatenbanken (CVE/NVD, OSV, EUVD) korreliert; Schwachstelleninformationen werden getrennt über VEX/CSAF kommuniziert und nicht in die statische Stückliste gemischt. Für

kritische Artefakte wird zusätzlich eine manipulationssichere Build-Provenance nach dem SLSA-Framework vorgehalten [SLSA]; Qualitätsanforderungen an Stücklisten enthält BSI TR-03183-2 [TR03183], den rechtlichen Rahmen der Cyber Resilience Act [CRA].

Der Wirksamkeitstest ist konkret: Bei einer neu bekannten Schwachstelle in einer verbreiteten Bibliothek muss allein anhand der Stücklisten in Minuten feststellbar sein, welche eigenen Artefakte betroffen sind. Von kritischen Softwarelieferanten wird die Bereitstellung aktueller Stücklisten vertraglich eingefordert; Institutionen ohne eigene Entwicklung nutzen das Prinzip vollständig über die Beschaffung.

Messung: SBOM-Coverage — Anteil der ausgelieferten Artefakte mit aktueller, automatisch erzeugter Stückliste (Zielwert: vollständige Abdeckung). Typische Fehler: Stücklisten werden erzeugt, aber nie ausgewertet; nur oberste Abhängigkeiten werden erfasst; die Stückliste wird nicht je Version fortgeschrieben und passt nicht zum ausgelieferten Stand.

MRIS.bd.1.M9 Integration automatisierter Sicherheitstests in die Entwicklungs-Pipeline (S)

In der Pipeline werden drei Prüfklassen kombiniert: statische Code-Analyse (SAST, z. B. Semgrep, SonarQube), dynamische Tests der laufenden Anwendung (DAST, z. B. OWASP ZAP) und Abhängigkeitsanalyse (SCA, z. B. Dependency-Track, Snyk), ergänzt um KI-gestütztes Code-Scanning. Funde der Stufen High oder Critical mit hoher Konfidenz blockieren die Übernahme in den Hauptzweig (Pre-Merge-Block); Schwachstellen aus dem Katalog aktiv ausgenutzter Lücken (KEV) blockieren sofort. Die laufenden Systeme werden täglich gescannt.

KI-generierter Code wird als eigene Risikokategorie behandelt: KI-Coding-Assistenten erzeugen regelmäßig unsichere Muster wie hartkodierte Zugangsdaten, fehlende Eingabe-Validierung, unsichere Defaults und veraltete Abhängigkeiten. Pre-Commit-Prüfungen mit Regelsätzen gegen diese KI-typischen Anti-Patterns und ein regelmäßiges Audit-Reporting mit Trendverfolgung adressieren das. Das Shift-Left-Prinzip ist die strukturelle Antwort auf den Kollaps des Patch-Gap-Fensters (siehe Baustein, Kap. 2.1): Es entzieht dem Angreifer das Zeitfenster zwischen Exploit-Verfügbarkeit und Produktions-Patch.

Messung: Zeit vom Bekanntwerden einer aktiv ausgenutzten Schwachstelle bis zum ausgerollten Patch (Richtwert für internet-exponierte Systeme: unter 24 Stunden) sowie Anteil der Pipelines mit Sicherheitsprüfung und Pre-Merge-Block. Typischer Fehler: Die Prüfungen laufen, aber Blockierungen werden routinemäßig per Ausnahme umgangen — die Kontrolle existiert nur auf dem Papier.

MRIS.bd.1.M10 Automatisierung der Vorfallsreaktion mit abgesicherten Playbooks (S)

Für eindeutige Vorfallstypen werden vordefinierte Playbooks auf einer SOAR-Plattform hinterlegt, die Containment-Aktionen automatisch einleiten; menschliche Eskalation erfolgt gezielt bei mehrdeutigen Fällen. KI-spezifische Playbooks decken mindestens ab: Massen-Datenexfiltration, parallele Multi-Vector-Angriffe, Lieferketten-Kompromittierung, Credential-Stuffing-Wellen und MFA-Fatigue-Kampagnen. Übungen für drei bis fünf parallele Vorfälle finden mindestens halbjährlich statt, da agentische Angreifer Parallelität gezielt als Überlastungstaktik einsetzen.

Die Automatisierungsschicht selbst wird als Angriffsfläche behandelt und gehärtet: signierte, authentisierte Auslöser, vollständiger Audit-Trail jeder Playbook-Ausführung mit Ausführungs-Identität, Rate-Limits pro Playbook, menschliche Freigabe (Human-in-the-Loop) für Aktionen mit großem Schadensradius (Massen-Kontosperrung, Netz-Isolation größerer Segmente, Zertifikats-Widerruf) und Schutz der Playbook-Definitionen wie Quellcode. Ohne diese Härtung kann die Reaktionsautomatik gegen die eigene Organisation gerichtet werden, etwa durch provozierte Fehlalarme.

Für Institutionen ohne eigenes Security-Team sind Managed-Detection-and-Response-Dienste mit 24/7-Betrieb und vertraglich vereinbarter Eindämmungszeit die gangbarere Alternative zum SOAR-Eigenbetrieb. Messung: Zeit bis zur Eindämmung eindeutiger Vorfälle (Richtwert: unter zehn Minuten bei hoher Eindeutigkeit). Typischer Fehler: Automatisierung wird eingeführt, bevor die Erkennung (M4) stabil ist — die Playbooks reagieren dann auf unzuverlässige Signale.

MRIS.bd.1.M11 Einrichtung einer kontinuierlichen automatisierten Wirksamkeitsprüfung (S)

Sicherheitsvorgaben werden, wo möglich, als ausführbare Regeln formalisiert (Policy-as-Code, z. B. mit OPA oder vergleichbaren Engines) und doppelt eingebunden: als Prüfung vor der Bereitstellung (Pre-Deploy-Checks in der CI) und als Durchsetzung im laufenden Betrieb. Die Einhaltung der Soll-Zustände wird kontinuierlich und mindestens täglich geprüft; Abweichungen lösen automatisch Vorgänge aus (z. B. Incident-Tickets) und werden in Dashboards mit quantifizierten Abweichungsindikatoren sichtbar.

Der Kern der Maßnahme ist die Verkürzung der Detektionslatenz für Kontrollabweichungen von Monaten (periodisches Audit) auf Minuten. Unter KI-Bedingungen ist das entscheidend, weil eine unbemerkt abgedriftete Konfiguration in der Zeit bis zum nächsten Audit bereits ausgenutzt sein kann. Die Maßnahme ersetzt periodische Audits nicht, verlagert deren Schwerpunkt aber auf die Prüfung der Regelqualität.

Messung: Continuous-Monitoring-Coverage — Anteil der im Statement of Applicability geführten Maßnahmen mit aktiver automatisierter Prüfung; Richtwerte: mindestens 60 % nach zwölf Monaten, mindestens 80 % nach 24 Monaten. Wesentliche Maßnahmen ohne automatisierte Prüfung sind nur befristet und begründet zulässig. Typischer Fehler: Es werden Dashboards aufgebaut, ohne dass Abweichungen eine verbindliche Reaktion auslösen.

MRIS.bd.1.M12 Einsatz signierter Container-Images und geschützter Ausführungsumgebungen (H)

Für Container gilt: Base-Images stammen ausschließlich aus kontrollierten Registries, werden signiert (z. B. Sigstore/cosign) und vor dem Deployment automatisiert gescannt. Die Durchsetzung erfolgt technisch über Admission-Controller-Policies der Plattform (z. B. OPA Gatekeeper, Kyverno), sodass unsignierte oder ungeprüfte Images gar nicht erst starten. Die Signatur schafft eine kryptografische Hartbarriere gegen manipulierte Images in der Lieferkette.

Für Workloads mit besonders hohem Schutzbedarf werden geschützte Ausführungsumgebungen eingesetzt (Trusted Execution Environments, z. B. Intel TDX, AMD SEV-SNP, ARM CCA), deren Echtheit über Remote-Attestation nachgewiesen wird. Confidential Computing ist derzeit die härteste verfügbare Barriere gegen Angreifer mit Zugriff auf die Infrastruktur des Betreibers und daher insbesondere für regulierte Workloads und hochsensible Datenverarbeitung relevant.

Messung: Anteil der produktiven Container aus signierten, geprüften Images (Zielwert: vollständig) sowie die Abdeckung von Confidential Computing bei den als hochsensibel eingestuften Workloads; Toleranzgrenze: kein produktiver Container ohne gültige Signatur und Prüfung. Typischer Fehler: Die Signaturprüfung ist im Admission-Controller im Audit-Modus konfiguriert und blockiert nicht.

MRIS.bd.1.M13 Umsetzung einer nachweisbaren Mandantentrennung (H)

Die Trennung wird auf drei Ebenen dokumentiert und umgesetzt: für Daten (mandantenspezifische Schlüssel, logische oder physische Trennung), für Netzwerke (mandantenspezifische VPCs, Namespaces, Software-Defined-Networking-Regeln) und für Rechenressourcen (mandantenspezifische Scheduling-Policies). Entscheidend ist der Übergang von der

konzeptionellen Behauptung zum Nachweis: Die Wirksamkeit der Trennung wird durch regelmäßige, möglichst automatisierte Trennungstests belegt.

Als Richtwerte für die Teststufen gelten: jährlich intern, je Major-Release automatisiert, ergänzend eine Prüfung durch unabhängige Dritte. Die Maßnahme begrenzt strukturell den Schadensradius: Ein Angreifer, der in einem Mandanten Fuß fasst, wird nachweislich daran gehindert, auf andere Mandanten zuzugreifen — unabhängig davon, wie schnell er operiert.

Messung: Ergebnis der Trennungstests (kein mandantenübergreifender Zugriff nachweisbar) sowie Häufigkeit und Abdeckung der Tests; Nachweis über Trennungskonzepte, Testprotokolle und Behebungsnachweise. Typischer Fehler: Die Trennung wird nur auf Anwendungsebene (Mandanten-ID im Code) umgesetzt — ein einziger Anwendungsfehler hebt sie dann vollständig auf.

MRIS.bd.1.M14 Durchführung bedrohungsgeleiteter Penetrationstests mit KI-Szenarien (H)

Threat-Led Penetration Tests werden mit externen Red-Teams durchgeführt, die reale, auf die aktuelle Bedrohungslage zugeschnittene Szenarien nachstellen. KI-spezifische Szenarien umfassen mindestens: KI-gestütztes Spear-Phishing einschließlich Deepfake-Voice, fragmentierte Angriffsketten über Mikroschritte, parallele Multi-Vector-Angriffe, Lieferketten-Kompromittierung und Cloud-API-Missbrauch über kompromittierte Service-Konten. Der Szenario-Pool wird multi-systemisch angelegt (mindestens zwei unabhängige agentische Systeme) und nicht an ein Einzelmodell gebunden. Kritische Funktionen werden mindestens jährlich getestet; zwischen den Zyklen finden Purple-Team-Übungen mit ATT&CK-Coverage-Zuordnung statt.

Vollwertige Tests nach TIBER-EU-Methodik verursachen typisch Kosten im mittleren sechsstelligen Bereich pro Übung. Für Institutionen außerhalb des DORA-Anwendungsbereichs sind kostengünstigere Formate ausreichend wirksam: Purple-Team-Übungen mit MITRE-ATT&CK-Szenarien, Open-Source-Adversary-Emulation (z. B. Caldera, Atomic Red Team) und Breach-and-Attack-Simulation-Plattformen.

Der Kern der Maßnahme: Getestet wird, ob die Schutzmaßnahmen gegen reale KI-Vorgehensweisen wirken — nicht, ob sie dokumentiert sind. Das reduziert das Risiko falsch-positiver Compliance-Zusicherungen. Messung: Closure-Rate — Anteil der festgestellten Schwachstellen, die innerhalb des vereinbarten Zeitfensters behoben werden (Richtwert: mindestens 90 %). Typischer Fehler: Testberichte werden abgelegt, ohne dass die Behebung nachgehalten wird.

3 Weiterführende Informationen

3.1 Wissenswertes

Herkunft und Zuordnung der Anforderungen. Die Anforderungen dieses Bausteins operationalisieren die dreizehn Mythos-Härtungs-Controls (MHC) des MRIS-Frameworks; A1 setzt zusätzlich die MRIS-Handlungsempfehlung zur Sofort-Neubewertung um [MRIS]. Für die Übernahme in das Statement of Applicability und den Abgleich mit ISO/IEC 27001 Annex A gilt folgende Zuordnung: A1 entspricht der MRIS-Neubewertung (Kap. 10.2, gesamter Annex A); A2 entspricht MHC-03 und flankiert A.5.17 sowie A.8.5; A3 entspricht MHC-08 und vertieft A.8.13; A4 entspricht MHC-05 und flankiert A.8.16 sowie A.8.12; A5 entspricht MHC-04 und flankiert A.5.16, A.8.20 sowie A.8.21; A6 entspricht MHC-01 und flankiert A.8.24; A7 entspricht MHC-13 und erweitert A.5.9, A.5.16 sowie A.8.27; A8 entspricht MHC-02 und flankiert A.5.21 sowie A.8.30; A9 entspricht MHC-09 und flankiert A.8.29 sowie A.8.25; A10 entspricht MHC-11 und ergänzt A.5.24 bis A.5.26; A11 entspricht MHC-10 und stützt A.5.35 sowie A.5.36; A12 entspricht MHC-06 und

flankiert A.5.23 sowie A.8.31; A13 entspricht MHC-07 und flankiert A.5.23 sowie A.8.22; A14 entspricht MHC-12 und ergänzt A.5.35 sowie A.8.29.

Umsetzungsreihenfolge. Die Zuordnung der Anforderungen zu Basis, Standard und erhöhtem Schutzbedarf folgt der MRIS-Standardpriorisierung P0/P1/P2 für typische Enterprise-Umgebungen. Sie ist ein Ausgangspunkt, kein universeller Plan: Bei langlebig vertraulichen Daten rückt A6 nach vorn, bei produktivem KI-Agenten-Einsatz A7; bei Container- oder Multi-Tenant-Plattformen gewinnen A12 und A13 an Dringlichkeit. Als Abhängigkeitsregeln gelten insbesondere: Neubewertung (A1) vor Investition in neue Maßnahmen; Erkennung (A4) vor Reaktionsautomatisierung (A10); Stücklisten (A8) als Grundlage für Pipeline-Testing (A9); Validierung durch Tests (A14) nach der Basisumsetzung [MRISIG].

Reifegrade. Der Implementation Guide beschreibt je Maßnahme einen kumulativen Reifegrad-Pfad in drei Stufen (Initial, Defined, Managed) mit Stage-Gates und Kennzahlen. Diese Stufen eignen sich für die interne Fortschrittsmessung und die Berichterstattung an die Leitungsebene [MRISIG].

Integration ins ISMS. Jede Anforderung dieses Bausteins ist so formuliert, dass sie in das Statement of Applicability eines bestehenden ISO-27001-zertifizierten ISMS übernommen werden kann. Nicht anwendbare Anforderungen werden dort begründet ausgeschlossen — typische Fälle: A8 und A9 bei Institutionen ohne eigene Softwareentwicklung (das Prinzip wirkt dann über die Beschaffung, siehe M8), A13 ohne mandantenfähigen Betrieb. Gesetzliche Meldepflichten aus NIS2 und DORA sind bewusst nicht als eigene Anforderung geführt: Sie sind direkte Rechtspflichten und werden im ISMS über die Vorfallbehandlung umgesetzt; A10 liefert die dafür nötige Reaktionsfähigkeit.

3.2 Quellenverweise

[MRIS] Mythos-resistente Informationssicherheit (MRIS) — Framework, Version 1.9 (DE), Richard Peddi, August 2026, CC BY-SA 4.0.

[MRISIG] MRIS Implementation Guide, Version 1.5 (DE), Richard Peddi, August 2026, CC BY-SA 4.0.

[ATTACK] MITRE ATT&CK Knowledge Base, The MITRE Corporation, laufend aktualisiert, https://attack.mitre.org, zuletzt aufgerufen am 14.08.2026.

[FIPS] FIPS 203 (ML-KEM), FIPS 204 (ML-DSA), FIPS 205 (SLH-DSA), NIST, August 2024.

[TR02102] Technische Richtlinie TR-02102: Kryptographische Verfahren — Empfehlungen und Schlüssellängen, BSI, jeweils aktuelle Fassung.

[SP180038] NIST SP 1800-38: Migration to Post-Quantum Cryptography, NIST NCCoE.

[TR03183] Technische Richtlinie TR-03183-2: Cyber-Resilienz-Anforderungen — Software Bill of Materials, BSI.

[SLSA] Supply-chain Levels for Software Artifacts (SLSA), OpenSSF, https://slsa.dev, zuletzt aufgerufen am 14.08.2026.

[CRA] Verordnung (EU) 2024/2847 über horizontale Cybersicherheitsanforderungen (Cyber Resilience Act), Annex I.