MRIS IMPLEMENTATION GUIDE

Umsetzungsleitlinie zu den dreizehn Mythos-Härtungs-Controls

Begleitdokument zu MRIS, Kapitel 9 (Version 1.6) — Mythos-resistente Informationssicherheit

ISO-27002-Tiefe · Evidenzbasiert · Praxisorientiert

Version 1.2 | Juni 2026
Ersteller: Richard Peddi

ZIELGRUPPE

CISOs und ISMS-Verantwortliche, die die dreizehn „Mythos-Härtungs-Controls“ aus MRIS, Kapitel 9 praktisch umsetzen wollen.

Zweck, Unabhängigkeit und Konventionen

Dieses Dokument ist der Umsetzungs-Layer zu MRIS, Kapitel 9. MRIS 1.6 listet die dreizehn Mythos-Härtungs-Controls (MHC) knapp in Annex-A-Logik; dieses Begleitdokument liefert zu jedem MHC die Umsetzungsleitlinie in ISO-27002-Logik — so, dass eine CISO-Funktion Zweck, organisatorische und technische Umsetzung, Reifegrad und Nachweisführung ohne zusätzliche Spezialliteratur nachvollziehen kann. Es ersetzt MRIS 1.6 nicht, sondern führt dessen Katalog praktisch aus. Die Quellenverweise dieses Leitfadens sind bereits auf den verifizierten aktuellen Stand (Juni 2026) gebracht.

Fixe Gliederung je MHC

- Kopf »Auf einen Blick«
- 1. Control-Statement
- 2. Zweck und Bedrohungsbezug
- 3. Organisatorische Umsetzung (CISO)
- 4. Technische Umsetzung (IT, schließt mit Wirksamkeitstest)
- 5. Umsetzungsbeispiele (Beispiel A / Beispiel B)
- 6. Reifegrad-Pfad
- 7. Messung und Audit-Nachweis
- 8. Typische Fehler
- 9. Abgrenzung, Restrisiko und Verweise.

Die in diesem Leitfaden je MHC enthaltenen Abschnitte „Messung und Audit-Nachweis“ sind bewusst schlank gehalten. Sie sollen eine praktikable Nachweisführung ermöglichen, ohne die Umsetzung durch übermäßige Dokumentationsanforderungen zu erschweren.

Organisationen mit erhöhtem Audit-, Regulatorik- oder Kundenanforderungsniveau können die Nachweisführung freiwillig erweitern. Dafür kann je anwendbarem MHC ein ergänzendes Audit-Raster verwendet werden:

- Control Owner: verantwortliche Rolle oder Organisationseinheit
- Scope: betroffene Systeme, Services, Standorte, Plattformen oder Datenklassen
- Applicability: Begründung, warum das MHC anwendbar oder nicht anwendbar ist
- Implementation Status: geplant, teilweise umgesetzt, umgesetzt, wirksam geprüft
- Evidence Type: Richtlinie, Konfiguration, Log, Report, Testprotokoll, Ticket, Risikoakzeptanz
- Test Frequency: Anlass und Häufigkeit der Wirksamkeitsprüfung
- Sampling Logic: Stichprobenlogik bei großen Systemlandschaften
- Exceptions: dokumentierte Abweichungen, Ausnahmen und Kompensationsmaßnahmen
- Risk Acceptance: akzeptierte Restrisiken inklusive Freigabeinstanz
- KPI/KRI: verwendete Kennzahlen und Zielwerte
- Last Effectiveness Review: Datum und Ergebnis der letzten Wirksamkeitsprüfung

Dieses erweiterte Raster ist kein verpflichtender Bestandteil des MRIS Implementation Guide. Es ist als optionaler Zusatz für Organisationen gedacht, die eine stärkere Audit-Tiefe benötigen, etwa aufgrund regulatorischer Anforderungen, Kundenprüfungen, Zertifizierungsvorbereitung oder eines höheren ISMS-Reifegrads. Für Organisationen mit geringerem Reifegrad genügt zunächst die je MHC beschriebene Mindestnachweisführung.

Lizenz und Haftungsausschluss

Haftungsausschluss

Dieser Leitfaden stellt eine technische und organisatorische Referenz dar. Es ersetzt keine individuelle Risikoanalyse, keine rechtliche Beratung und keine auditspezifische Prüfung. Die hier vorgenommene Bewertung einzelner Controls bezieht sich auf den zum Erstellungszeitpunkt öffentlich dokumentierten Stand agentischer KI-Bedrohungen. Die Bewertung kann sich bei neuer Evidenz verschieben.

Zitierempfehlung

Peddi, Richard (2026): MRIS IMPLEMENTATION GUIDE, Version 1.2.

Inhalt

Priorisierung der MHC — Threat-Korrelation und Roadmap

Die dreizehn MHC sind nicht als lineare Reihenfolge von MHC-01 bis MHC-13 zu verstehen. Ohne Priorisierung wirkt jedes Control gleich dringlich — und genau das erschwert die Umsetzungsentscheidung. Priorisiert wird nicht das Control an sich, sondern sein Beitrag zur Reduktion der für die Organisation relevantesten Bedrohungen. Dieses Kapitel verbindet dafür zwei sich ergänzende Sichten: die messbare Wirkungsbreite jedes MHC (wie viele degradierte Controls es trägt) und die bedrohungs- und aufwandsbezogene Reihenfolge (Threat-Korrelation). Die Priorisierung bleibt anschlussfähig an das übergeordnete MRIS-Prinzip: strukturelle Controls vor neuen Detection- oder Automatisierungs-Capabilities.

Auf einen Blick

- Zweck: nachvollziehbare, bedrohungsorientierte Umsetzungs-Roadmap statt linearer MHC-Reihenfolge.
- Grundlage: Wirkungsbreite (Coverage) je MHC plus Threat Priority Score.
- Ergebnis: Standard-Tiering P0/P1/P2 als anpassbares Template, mit Abhängigkeitsregeln.
- Wichtig: MHC-01 und MHC-13 haben Coverage 0, können aber je nach Architektur und Datenlage P0 sein.

1. Wirkungsbreite: wie viele degradierte Controls jedes MHC trägt

Die folgende Übersicht zeigt, wie viele der unter Mythos degradierten Controls (Kategorien „teilweise degradiert" und „reine Reibung", zusammen 41 Controls) jedes MHC flankiert. Die Zahlen sind gegen die Bewertungsmatrix in MRIS, Anhang A, geprüft (Gesamtverteilung 29 standfest / 37 teilweise degradiert / 4 reine Reibung / 23 nicht betroffen). Da ein degradiertes Control von mehreren MHC flankiert werden kann, ist die Spaltensumme (44) größer als 41. Die Spalte „auch standfest" nennt zusätzlich flankierte, bereits standfeste Controls — mehrere MHC härten also über die degradierten Controls hinaus auch robuste Controls weiter.

MHC	Degradierte Controls	Auch standfest flankiert	Charakter
MHC-05 Verhaltensbasierte Detection	10	A.8.15	breit querschnittlich
MHC-02 SBOM/Build-Provenance	6	—	Lieferkette
MHC-03 Phishing-resistente MFA	5	—	Identität/Auth
MHC-04 Workload-Identität/Zero Trust	5	A.5.16, A.8.2	Identität/Zugriff
MHC-09 AI-Security-Testing	5	A.8.29	sichere Entwicklung
MHC-11 SOAR/Tier-1-Automation	4	—	Incident Response
MHC-10 Continuous Control Monitoring	3	A.8.9	Prüfung
MHC-07 Multi-Tenancy-Isolation	2	—	Cloud
MHC-08 Unveränderliche Backups	2	A.5.30, A.8.13, A.8.14	Resilienz
MHC-06 Container/Confidential Computing	1	A.8.31	eng
MHC-12 Threat-Led Pentest	1	A.8.29	eng, validierend
MHC-01 Post-Quantum	0	A.8.24	flankiert nur Standfest
MHC-13 AI-Agent-Governance	0	A.5.9, A.5.16, A.8.27	flankiert nur Standfest

MHC-05 ist mit Abstand das breiteste Querschnitts-Control. Wirkung ist messbar ungleich verteilt — das ist das stärkste Argument gegen die Wahrnehmung, jedes MHC sei gleich wichtig.

2. Zwei Prioritäts-Sichten und ihr Zusammenhang

- Wirkungsbreite (Coverage): unbedingte Hebelwirkung — wie viele degradierte Controls ein MHC trägt. Hohe Coverage bedeutet hohen Hebel über viele Risiken zugleich.
- Bedrohungs- und Aufwandsbezug: die Reihenfolge — abhängig von Bedrohungslage, Auswirkung, Exposition, Control-Lücke, Abhängigkeiten und Aufwand.

Beide Sichten widersprechen sich nicht, sie ergänzen sich. Coverage allein würde MHC-01 und MHC-13 ans Ende stellen, weil sie nur standfeste Controls flankieren. Das wäre falsch: MHC-13 ist für jede Organisation mit AI-Agenten potenziell P0 (es adressiert den Kern der Mythos-Bedrohungslage), MHC-01 für langlebig vertrauliche Daten. Daher gilt: nach Hebelwirkung ordnen, aber mit einem Bedrohungs- und Anwendbarkeits-Gate — nie allein nach Coverage. „Low hanging fruits" sind dann Controls mit hoher Hebelwirkung, großer Control-Lücke und geringem Aufwand, unter Beachtung der Abhängigkeiten.

3. Bewertungslogik: Threat Priority Score

Für die Reihenfolge wird je Bedrohungsszenario ein einfacher, transparenter Score gebildet. Er macht sichtbar, ob ein MHC wegen hoher Bedrohungslage, hoher Auswirkung, starker Exposition oder großer Control-Lücke priorisiert wird.

Threat Priority Score = Threat-Relevanz × Business Impact × Exposure × Control Gap

Faktor	Leitfrage	Skala
Threat-Relevanz	Wie realistisch, aktuell und relevant ist das Angriffsszenario für die Organisation?	1 = gering … 5 = sehr hoch
Business Impact	Wie schwer wären Folgen für DORA-/BIA-kritische Services, Kundendaten, Verfügbarkeit, Reputation oder Finanzen?	1 = gering … 5 = existenziell
Exposure	Wie stark ist die Organisation exponiert (Internet-facing, Cloud, Lieferanten, Admin-Zugänge, externe Nutzer)?	1 = gering … 5 = stark exponiert
Control Gap	Wie groß ist die Lücke gegenüber dem MHC-Zielbild?	1 = gut abgedeckt … 5 = kaum abgedeckt

Beispiel: KI-gestütztes Phishing mit Threat-Relevanz 5, Business Impact 4, Exposure 5 und Control Gap 4 ergibt einen Score von 400 — sehr hohe Priorität, vorrangig für MHC-03, flankiert durch MHC-05 und MHC-11.

Der Threat Priority Score ist ein qualitativer Priorisierungsindex, kein Ersatz für eine formale Risikoanalyse (ISO/IEC 27005) oder eine organisationsspezifische Risikomatrix. Die Multiplikation der Faktoren macht relative Handlungsprioritäten sichtbar; der Wert ist ordinal, nicht metrisch — ein Score von 400 ist nicht „doppelt so dringlich“ wie 200, sondern dient der Reihung und Triage.

4. Threat-to-MHC-Matrix

Die Matrix zeigt, welche MHC gegen welche Bedrohung besonders wirksam sind. Sie eignet sich als Basis für Statement of Applicability, Risk Treatment Plan und Umsetzungs-Roadmap.

Bedrohung / Angriffsszenario	Primäre MHC	Begründung	Priorität
KI-gestütztes Phishing / Credential Theft	MHC-03, MHC-05, MHC-11	MHC-03 verhindert Credential-Missbrauch, MHC-05 erkennt auffällige Nutzung, MHC-11 automatisiert die erste Reaktion.	sehr hoch
Lateral Movement nach Erstkompromittierung	MHC-04, MHC-05, MHC-11, MHC-12	MHC-04 begrenzt seitliche Bewegung über Workload-Identität, MHC-05 erkennt Angriffsketten, MHC-12 validiert die Wirksamkeit.	sehr hoch
Software-Supply-Chain-Angriff	MHC-02, MHC-09, MHC-06	SBOM, Build-Provenance, Pipeline-Testing und Image-Signatur greifen ineinander.	hoch
Ransomware / destruktiver Angriff	MHC-08, MHC-05, MHC-11, MHC-12	Unveränderliche Backups sichern die Wiederherstellung; Detection und SOAR verkürzen die Reaktionszeit; TLPT/Purple Team prüft die Resilienz.	sehr hoch
Manipuliertes Container-Image	MHC-06, MHC-02, MHC-09	Signierte Images und Admission Control benötigen Herkunfts- und Schwachstelleninformationen aus den Supply-Chain-Controls.	mittel bis hoch
Tenant Breakout / mandantenübergreifender Zugriff	MHC-07, MHC-04, MHC-06	Relevant bei SaaS-/Multi-Tenant-Plattformen; die Trennung muss technisch durchgesetzt und getestet werden.	hoch, falls anwendbar
AI-Agent-Missbrauch / Prompt Injection / Tool Abuse	MHC-13, MHC-04, MHC-10, MHC-09	Agenten brauchen eigene Identitäten, begrenzte Tool-Rechte, Policy-Kontrolle und eine sichere Pipeline.	hoch, falls AI-Agenten im Einsatz
Harvest now, decrypt later / Quantenrisiko	MHC-01	Besonders relevant für langlebig vertrauliche Daten und Krypto-Agilität.	mittel bis hoch
Unbemerkte Control-Abweichungen	MHC-10	Continuous Control Monitoring erkennt Abweichungen laufend statt nur punktuell im Audit.	mittel
Falsch-positive Sicherheitsannahmen	MHC-12	Threat-led Tests zeigen, ob dokumentierte Controls tatsächlich wirken.	hoch nach Basisumsetzung

5. Abhängigkeiten als Roadmap-Regeln

Die folgende Logik verhindert, dass Controls isoliert umgesetzt werden, obwohl sie erst durch andere MHC wirksam oder sinnvoll prüfbar werden.

Abhängigkeit	Bedeutung
MHC-02 → MHC-09	AI-gestütztes Security-Testing in der Pipeline braucht eine Build-/CI-Basis und profitiert direkt von SBOM und Komponenteninformationen.
MHC-04 → MHC-13	AI-Agent-Governance braucht technische Identitäten, capability-scoped Zugriff und klare Trennung von persönlichen Benutzerkonten.
Logging (A.8.15) → MHC-05 → MHC-11	SOAR-Automation ist nur belastbar, wenn Detection auf zuverlässigen, zentralen und manipulationssicheren Logs basiert.

Abhängigkeit	Bedeutung
MHC-05 + MHC-11 → MHC-12	Threat-led Tests validieren, ob Detection und Response in realistischen Angriffsketten funktionieren.
MHC-02 + MHC-04 → MHC-06	Container-Sicherheit wird stärker, wenn Herkunft der Images und Workload-Identitäten sauber kontrolliert sind.
MHC-04 + MHC-06 → MHC-07	Multi-Tenancy-Isolation nutzt Identitäts-, Netzwerk-, Ressourcen- und Laufzeittrennung.

6. Standard-Priorisierung P0 / P1 / P2

Das folgende Tiering ist ein Beispiel für typische Enterprise-Umgebungen. Es ist mit der tatsächlichen Bedrohungslage, der Asset-Kritikalität und der Anwendbarkeit abzugleichen. MHC-01 und MHC-13 stehen formal in P2, rücken aber bei langlebig vertraulichen Daten (MHC-01) bzw. produktivem AI-Agenten-Einsatz (MHC-13) nach oben. Das Standard-Tiering ist kein universeller Umsetzungsplan, sondern ein Ausgangspunkt; es ist je Organisation an Exposition, Architektur, Asset-Kritikalität, bestehende Controls, Ressourcenlage und regulatorischen Kontext anzupassen.

Priorität	MHC	Control	Begründung
P0 — sofort	MHC-03	Phishing-resistente MFA	Hohe Eintrittswahrscheinlichkeit, schnelle Risikoreduktion bei Admins und externen Zugängen.
P0 — sofort	MHC-08	Unveränderliche Backups und Recovery-Validierung	Basisfähigkeit gegen Ransomware und destruktive Angriffe.
P0 — sofort	MHC-05	Verhaltensbasierte Detection und Kill-Chain-Korrelation	Ohne Detection keine belastbare Reaktion und keine sinnvolle SOAR-Automation.
P0 — sofort	MHC-04	Workload-Identität und Zero Trust	Reduziert Lateral Movement und ist Enabler für AI-Agenten.
P1 — danach	MHC-02	SBOM und Build-Provenance	Grundlage für Supply-Chain-Transparenz und Pipeline-Testing.
P1 — danach	MHC-09	AI-gestütztes Security-Testing	Wirksam bei vorhandener Pipeline und SBOM-Basis.
P1 — danach	MHC-11	SOAR Tier-1-Automation	Erst automatisieren, wenn Detection stabil ist.
P1 — danach	MHC-10	Continuous Control Monitoring	Querschnitts-Control für laufende Abweichungserkennung.
P2 — risikobasiert	MHC-06	Container-Sicherheit und Confidential Computing	Hoch relevant bei Container-/Cloud-Plattformen.
P2 — risikobasiert	MHC-07	Multi-Tenancy-Isolation	Nur bei Multi-Tenant-Plattformen oder SaaS-Betrieb direkt anwendbar.
P2 — risikobasiert	MHC-13	AI-Agent-Governance	Hoch relevant bei AI-Agenten, Tool-Calling oder Automatisierung; dann nach P0 hochstufen.
P2 — risikobasiert	MHC-01	Post-Quantum-Strategie	Höher priorisieren bei langlebig vertraulichen Daten.
P2 — risikobasiert	MHC-12	Threat-Led Penetration Testing	Validierung nach Basisumsetzung, danach regelmäßig.

7. Vorgehen in vier Schritten

- Schritt 1 — Threat Landscape erstellen: relevante Angreifertypen, TTPs und Angriffsszenarien identifizieren; branchenspezifische Bedrohungen berücksichtigen (Finanzdienstleister, kritische Services, Lieferkette, Cloud, AI); Bedrohungen nach Aktualität, Eintrittswahrscheinlichkeit und Nähe zur Organisation bewerten.
- Schritt 2 — Threats mit MHC mappen: je Bedrohung festlegen, welche MHC präventiv, detektiv, reaktiv oder validierend wirken; Controls mit mehreren Bedrohungsbezügen höher gewichten; nicht anwendbare Controls im Statement of Applicability sauber begründen.
- Schritt 3 — Score berechnen: Threat-Relevanz, Business Impact, Exposure und Control Gap je 1 bis 5 bewerten; Score bilden und Aufwand sowie Abhängigkeiten berücksichtigen; bei gleichem Risiko zuerst Quick Wins mit hoher Risikoreduktion und geringer Abhängigkeit.
- Schritt 4 — Roadmap ableiten: P0 (Sofortmaßnahmen gegen hohe Bedrohungen und große Lücken), P1 (Controls mit Abhängigkeiten oder technischer Basisarbeit), P2 (kontextabhängige Controls und Validierung); regelmäßig neu bewerten, sobald sich Bedrohungslage, Architektur oder Regulatorik ändern.

8. Formulierung für Governance-Unterlagen

Zur risikobasierten Priorisierung der Mythos-Härtungs-Controls werden die MHC mit relevanten Bedrohungsszenarien korreliert. Je Bedrohung wird bewertet, welche Controls präventiv, detektiv, reaktiv oder validierend wirken. Die Priorisierung ergibt sich aus Threat-Relevanz, Business Impact, Exposition, bestehender Control-Lücke, Abhängigkeiten und Umsetzungsaufwand. So entsteht keine lineare MHC-Reihenfolge, sondern eine bedrohungsorientierte Roadmap, die für jedes Control begründet, warum es zu seinem Zeitpunkt umgesetzt wird.

MHC-01 — Post-Quantum-Strategie und kryptografisches Inventar

Auf einen Blick

- Operativer Nutzen: Schützt langlebig vertrauliche Daten vor dem Muster „heute abgreifen, später entschlüsseln" und schafft die Fähigkeit, kryptografische Verfahren bei Bedarf zügig auszutauschen.
- Betroffene Richtlinie: Kryptografierichtlinie (Verschlüsselung und Schlüsselverwaltung).
- Abhängigkeiten: keine Vorbedingung; setzt ein Verständnis der eingesetzten Verschlüsselung voraus (Inventar).
- Aufwandstreiber: hängt von der Zahl der Systeme und Datenflüsse mit Verschlüsselung sowie von der Austauschbarkeit der verwendeten Bibliotheken ab. Der konkrete Aufwand richtet sich nach den Ressourcen der Organisation.
- Primäre Kennzahl: Anteil der eingesetzten kryptografischen Verfahren, die im zentralen Inventar erfasst und nach Migrationspriorität eingestuft sind.
- Berührte Standards (ohne Erfüllungsanspruch): ISO/IEC 27002 A.8.24; NIST FIPS 203/204/205; BSI TR-02102; C5:2026 CRY-01; EU-Empfehlung zur PQC-Migration.

1. Control-Statement

Die Organisation führt ein Inventar aller eingesetzten kryptografischen Verfahren und verfolgt eine dokumentierte Strategie, um rechtzeitig auf quantensichere Verfahren umzustellen. Im Übergang werden hybride Verfahren eingesetzt.

2. Zweck und Bedrohungsbezug

Künftige Quantencomputer werden heute verbreitete Verschlüsselung (etwa RSA und elliptische Kurven) brechen können. Angreifer fangen daher bereits heute verschlüsselte Daten ab, um sie später zu entschlüsseln — das Muster „heute abgreifen, später entschlüsseln". Besonders betroffen sind Daten, die über viele Jahre vertraulich bleiben müssen. KI beschleunigt zudem das Auffinden und Ausnutzen schwacher oder veralteter Krypto-Implementierungen. Schutzziel ist, langlebig vertrauliche Daten rechtzeitig auf quantensichere Verfahren umzustellen und die Fähigkeit zu schaffen, Verfahren bei Bedarf zügig zu wechseln.

3. Organisatorische Umsetzung (CISO-Perspektive)

- Richtlinie: Die Kryptografierichtlinie wird um die Pflicht eines zentralen Krypto-Inventars und um eine Migrationsstrategie zu quantensicheren Verfahren ergänzt.
- Verantwortlichkeiten: Eine Stelle wird für die Pflege des Inventars und die Fortschreibung der Strategie benannt und im Statement of Applicability zugewiesen (Vorlage MRIS Anhang H).
- Risikoanalyse und SoA: Die Migrationsreihenfolge wird aus der Risikobewertung abgeleitet — Daten mit langer Vertraulichkeitsdauer zuerst. Behandeltes Risiko: spätere Entschlüsselung heute abgeflossener Daten (Anbindung an die Risikobewertungs-Brücke, MRIS Anhang F).
- Prozesse: regelmäßige (jährliche oder anlassbezogene) Überprüfung von Inventar und Strategie; Beobachtung der einschlägigen Standards; festgelegte Auslöser, Mittel und Übergangswege für die Umstellung.
- Beschaffung: bei neuen Systemen und Verträgen wird die Unterstützung quantensicherer bzw. austauschbarer Verfahren als Anforderung aufgenommen.

4. Technische Umsetzung (für IT-Fachleute)

- Krypto-Inventar: Erfassung aller eingesetzten Algorithmen, Schlüssellängen, Protokolle, Zertifikate und der zugrunde liegenden Bibliotheken; Priorisierung nach Schutzbedarf und Vertraulichkeitsdauer.
- Quantensichere Verfahren: 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. HQC steht als ergänzender Schlüsselaustausch in Standardisierung (Backup), FN-DSA (FALCON) als weitere Signatur in Vorbereitung.
- Hybride Verfahren: im Übergang werden klassische und quantensichere Verfahren kombiniert (z. B. im TLS-Schlüsselaustausch), sodass die Verbindung sicher bleibt, solange mindestens eines der beiden Verfahren hält.
- Crypto-Agility: Verschlüsselung wird so gekapselt, dass Verfahren ohne tiefe Eingriffe in die Anwendungen ausgetauscht werden können (zentrale Krypto-Bibliotheken, konfigurierbare Algorithmen).
- Vertiefung: Migrationsvorgehen in NIST SP 1800-38 (Migration to Post-Quantum Cryptography, NCCoE); Verfahrens- und Schlüssellängen-Empfehlungen in BSI TR-02102; Algorithmenspezifikation in FIPS 203/204/205.
- Wirksamkeitstest: Für ein priorisiertes System wird geprüft, dass das tatsächlich verwendete Verfahren dem Inventar und der Zielvorgabe entspricht (z. B. der ausgehandelte Schlüsselaustausch im TLS-Handshake) und dass ein Verfahren ohne Code-Änderung umgestellt werden kann.

5. Umsetzungsbeispiele

Beispiel A — Entwicklungs-/Plattform-Kontext. Ein Software-Anbieter weiß zunächst nicht genau, welche Verschlüsselung in seinen Diensten und Bibliotheken überhaupt zum Einsatz kommt. Er erstellt deshalb zuerst ein zentrales Verzeichnis aller verwendeten Verfahren, Schlüssel und Zertifikate und hält zu jedem Datenbestand fest, wie lange er vertraulich bleiben muss. Für die Verbindungen, über die besonders langlebige Daten fließen, stellt er den Schlüsselaustausch auf ein kombiniertes Verfahren um, das klassische und quantensichere Kryptografie zugleich nutzt. So bleiben diese Daten auch dann geschützt, wenn das klassische Verfahren später durch Quantencomputer gebrochen wird; zugleich kann das Verfahren bei Bedarf zentral gewechselt werden.

Beispiel B — Allgemeine Abteilung. Eine Organisation ohne eigene Entwicklung nutzt vor allem fertige Software und Cloud-Dienste. Sie kann die Verschlüsselung nicht selbst umstellen, verschafft sich aber einen Überblick, wo besonders langlebig vertrauliche Daten liegen (etwa Personal- oder Vertragsunterlagen), und fragt bei ihren Anbietern deren Fahrplan zur quantensicheren Umstellung ab. In neue Verträge nimmt sie die Unterstützung quantensicherer Verfahren als Anforderung auf. So steuert sie das Thema über Überblick und Beschaffung, auch ohne eigene technische Umsetzung.

6. Reifegrad-Pfad (kumulativ)

- Initial: kryptografisches Inventar dokumentiert.
- Defined: Migrationsstrategie verabschiedet; hybride Verfahren in einem Pilotbereich erprobt.
- Managed: hybride bzw. quantensichere Verfahren produktiv im Einsatz; jährliche Überprüfung automatisiert.

7. Messung und Audit-Nachweis

- Kennzahl: Anteil der eingesetzten Verfahren, die im Inventar erfasst und nach Migrationspriorität eingestuft sind (Zielwert: vollständige Erfassung der schützenswerten Bereiche).
- Nachweis: Krypto-Inventar, Migrationsstrategie mit Auslösern und Zeitplan, Pilot- und Produktivnachweise hybrider Verfahren.
- Prüflogik nach ISO/IEC 27005: dokumentiert? — Inventar und Strategie; verantwortlich? — benannte Stelle; Häufigkeit? — jährliche oder anlassbezogene Überprüfung; Toleranz? — keine langlebig vertraulichen Daten ohne Migrationsplan.

8. Typische Fehler

- Es wird eine Strategie formuliert, ohne dass ein vollständiges Inventar vorliegt — die wichtigsten Stellen bleiben unbekannt.
- Hybride Verfahren werden flächendeckend eingeführt, ohne nach Vertraulichkeitsdauer zu priorisieren — viel Aufwand ohne erkennbaren Nutzenschwerpunkt.
- Verschlüsselung ist fest im Code verankert, sodass ein späterer Wechsel des Verfahrens jeweils Eingriffe in die Anwendung erfordert.
- Die Umstellung wird allein als Zukunftsthema behandelt, obwohl der Datenabfluss von heute bereits das morgige Risiko bestimmt.

9. Abgrenzung, Restrisiko und Verweise

- Abgrenzung: betrifft die eingesetzte Kryptografie; die Verwaltung der Schlüssel selbst bleibt Teil der allgemeinen Kryptografie-Praxis (A.8.24).

- Restrisiko: solange klassische Verfahren parallel laufen, bleibt das Risiko, dass bereits abgeflossene Daten später entschlüsselt werden; quantensichere Verfahren sind zudem vergleichsweise jung und werden weiter beobachtet.
- ISO/IEC 27002:2022: A.8.24 Verwendung von Kryptographie.
- Framework: NIST FIPS 203 (ML-KEM), FIPS 204 (ML-DSA), FIPS 205 (SLH-DSA); NIST SP 1800-38 (Migrationsleitfaden, NCCoE); BSI TR-02102; C5:2026 CRY-01; EU-Empfehlung zur koordinierten PQC-Migration.

MHC-02 — SBOM und Build-Provenance

Auf einen Blick

- Operativer Nutzen: Macht sofort sichtbar, welche fremden Komponenten in der eigenen Software stecken, sodass bei einer neuen Schwachstelle in Minuten statt Tagen klar ist, ob und wo man betroffen ist.
- Betroffene Richtlinie: Richtlinie zur sicheren Softwareentwicklung und zum Lieferantenmanagement.
- Abhängigkeiten: keine Vorbedingung; setzt eine automatisierte Build-/CI-Pipeline voraus, um den Nutzen voll zu heben.
- Aufwandstreiber: hängt von Zahl und Aufbau der Build-Pipelines und der Tiefe der Lieferkette ab. Der konkrete Aufwand richtet sich nach den Ressourcen der Organisation.
- Primäre Kennzahl: Anteil der ausgelieferten Artefakte mit aktueller, automatisch erzeugter Stückliste (SBOM-Coverage).
- Berührte Standards (ohne Erfüllungsanspruch): ISO/IEC 27002 A.5.21/A.8.30; CRA Annex I; BSI TR-03183-2; C5:2026 DEV-13; SLSA.

1. Control-Statement

Für jede selbst erstellte und weitergegebene Software wird automatisch eine Stückliste der enthaltenen Komponenten (SBOM) erzeugt und gegen Schwachstellendatenbanken abgeglichen. Für kritische Artefakte wird zusätzlich nachvollziehbar belegt, wie und woraus sie gebaut wurden (Build-Provenance).

2. Zweck und Bedrohungsbezug

Moderne Software besteht zum großen Teil aus fremden Bausteinen (Open-Source-Bibliotheken, Frameworks). Wird in einem dieser Bausteine eine Schwachstelle bekannt — oder schleust ein Angreifer gezielt Schadcode in die Lieferkette ein —, ist ohne Übersicht oft tagelang unklar, ob und wo man betroffen ist. KI verkürzt die Zeit zwischen Bekanntwerden einer Schwachstelle und einem funktionierenden Angriff drastisch. Eine aktuelle Stückliste macht die eigene Betroffenheit sofort prüfbar; ein Nachweis über Herkunft und Bauweise erschwert es, manipulierte Bausteine unbemerkt unterzuschieben. Schutzziel ist, Angriffe über die Software-Lieferkette schnell zu erkennen und einzugrenzen.

3. Organisatorische Umsetzung (CISO-Perspektive)

- Richtlinie: Die Entwicklungs- und Lieferantenrichtlinie schreibt die automatische SBOM-Erzeugung für eigene und weitergegebene Artefakte vor sowie die vertragliche SBOM-Anforderung an kritische Softwarelieferanten.
- Verantwortlichkeiten: Im Statement of Applicability wird die Maßnahme einer Stelle zugewiesen (Vorlage MRIS Anhang H); Entwicklung und Einkauf wirken zusammen.
- Risikoanalyse und SoA: kritische Artefakte und Lieferanten werden aus der Risikobewertung priorisiert. Behandeltes Risiko: unbemerkte Schwachstellen oder

Manipulationen in fremden Komponenten (Anbindung an die Risikobewertungs-Brücke, MRIS Anhang F).
- Prozesse: fester Ablauf, wie bei einer neu bekannten Schwachstelle anhand der Stücklisten die Betroffenheit ermittelt und behoben wird; Aufbewahrung der Stücklisten je Softwareversion.
- Beschaffung: SBOM und, wo möglich, Provenance-Nachweise werden als Anforderung in Verträge mit kritischen Lieferanten aufgenommen.

4. Technische Umsetzung (für IT-Fachleute)

- SBOM-Erzeugung: automatisch in der Build-/CI-Pipeline, in einem etablierten, maschinenlesbaren Format — CycloneDX (als ECMA-424 standardisiert) oder SPDX (als ISO/IEC 5962 standardisiert). Erfasst werden auch die mittelbaren Abhängigkeiten (transitiv), nicht nur die obersten.
- Je Version eine SBOM: bei jeder Änderung einer Komponente wird eine neue Version mit eigener Stückliste erzeugt; Stücklisten werden versioniert aufbewahrt.
- Schwachstellen-Abgleich: automatisierte Korrelation der Komponenten gegen Schwachstellendatenbanken (CVE/NVD, OSV, EUVD der ENISA). Schwachstellendaten gehören nicht in die SBOM selbst, sondern werden getrennt über VEX bzw. CSAF kommuniziert.
- Build-Provenance: für kritische Artefakte wird nachvollziehbar und manipulationssicher festgehalten, aus welchen Quellen und mit welchem Build-Prozess sie entstanden sind (SLSA-Build-Provenance, höhere Stufe für kritische Artefakte).
- Werkzeuge (unverbindliche Beispiele): Erzeugung mit Syft, cdxgen oder Trivy; zentrale Verwaltung und Schwachstellen-Korrelation z. B. mit Dependency-Track.
- Vertiefung: Build-Provenance-Stufen im SLSA-Framework; SBOM-Qualitätsanforderungen in BSI TR-03183-2; rechtlicher Rahmen in CRA Annex I.
- Wirksamkeitstest: Für eine real bekannte Schwachstelle in einer verbreiteten Bibliothek (z. B. eine Log4j-artige Lücke) wird allein anhand der Stücklisten geprüft, ob und welche eigenen Artefakte die betroffene Komponente enthalten — das Ergebnis muss in Minuten vorliegen.

5. Umsetzungsbeispiele

Beispiel A — Entwicklungs-/Plattform-Kontext. Ein Software-Anbieter erfährt aus den Nachrichten von einer schweren Schwachstelle in einer weit verbreiteten Bibliothek. Bisher muss er mühsam alle Projekte einzeln durchsuchen, was Tage dauert. Er ergänzt deshalb seine Build-Pipeline so, dass bei jedem Bau automatisch eine vollständige Stückliste aller enthaltenen Bausteine entsteht und laufend gegen Schwachstellendatenbanken abgeglichen wird. Beim nächsten solchen Vorfall genügt eine Abfrage über die Stücklisten, um in Minuten zu sehen, welche Produkte die betroffene Komponente enthalten. Für seine wichtigsten Artefakte hinterlegt er zusätzlich einen manipulationssicheren Nachweis, woraus und wie sie gebaut wurden.

Beispiel B — Allgemeine Abteilung. Eine Organisation ohne eigene Entwicklung kann keine SBOM selbst erzeugen, ist aber auf zugekaufte Software angewiesen. Sie nimmt deshalb in ihre Verträge mit den wichtigsten Softwarelieferanten die Anforderung auf, zu jeder Lieferung eine aktuelle Stückliste bereitzustellen. Bei einer neu bekannten Schwachstelle kann sie so beim Hersteller gezielt nachhalten, ob das eingesetzte Produkt betroffen ist, statt auf vage Sammelmeldungen zu warten. So nutzt sie das Prinzip über die Beschaffung, auch ohne eigene Pipeline.

6. Reifegrad-Pfad (kumulativ)

- Initial: SBOM-Erzeugung in einzelnen Pipelines.
- Defined: SBOM für alle eigenen Artefakte; automatisierter Schwachstellen-Abgleich.
- Managed: SBOM auch von kritischen Lieferanten; Build-Provenance (SLSA) für kritische Artefakte; automatisierte Prüfung der Herkunftsnachweise.

7. Messung und Audit-Nachweis

- Kennzahl: SBOM-Coverage — Anteil der ausgelieferten Artefakte mit aktueller, automatisch erzeugter Stückliste (Zielwert: vollständige Abdeckung der eigenen Artefakte).
- Nachweis: SBOM-Dateien je Version, Konfiguration der Pipeline, Berichte des Schwachstellen-Abgleichs, Provenance-Nachweise kritischer Artefakte.
- Prüflogik nach ISO/IEC 27005: dokumentiert? — Pipeline-Konfiguration und Stücklisten; verantwortlich? — Entwicklungs-/Produktverantwortliche; Häufigkeit? — bei jedem Build, Abgleich fortlaufend; Toleranz? — kein ausgeliefertes Artefakt ohne aktuelle Stückliste.

8. Typische Fehler

- Stücklisten werden erzeugt, aber nie ausgewertet — im Schwachstellenfall liegt zwar eine SBOM vor, doch niemand gleicht sie ab.
- Nur die obersten Abhängigkeiten werden erfasst; gerade die tief verschachtelten Bausteine, in denen Schwachstellen stecken, fehlen.
- Die SBOM wird einmalig erstellt und nicht je Version fortgeschrieben, sodass sie nicht zum tatsächlich ausgelieferten Stand passt.
- Schwachstellendaten werden in die SBOM gemischt; dadurch wird die statische Stückliste mit dynamischen Informationen vermengt und schnell unbrauchbar.

9. Abgrenzung, Restrisiko und Verweise

- Abgrenzung: betrifft Transparenz und Herkunft der Software-Bausteine; das eigentliche Schließen der Schwachstellen ist Teil des Schwachstellen- und Patch-Managements, das automatisierte Testen behandelt MHC-09.
- Restrisiko: eine Stückliste erkennt bekannte, nicht aber unbekannte Schwachstellen; korrekt deklarierte, aber manipulierte Komponenten werden allein durch die SBOM nicht entlarvt — dafür dient die Provenance.
- ISO/IEC 27002:2022: A.5.19 Informationssicherheit in Lieferantenbeziehungen; A.5.20 Behandlung von Informationssicherheit in Lieferantenvereinbarungen; A.5.21 Umgang mit der Informationssicherheit in der IKT-Lieferkette; A.8.4 Zugriff auf den Quellcode; A.8.30 Ausgegliederte Entwicklung.
- Framework: C5:2026 DEV-13; CRA Annex I (SBOM-Pflicht; Schwachstellen-Meldepflichten ab 2026, Hauptpflichten für nach dem 11.12.2027 in Verkehr gebrachte Produkte); BSI TR-03183-2; DORA Art. 28; NIS2-DVO Nr. 5; SLSA-Framework; Formate CycloneDX (ECMA-424) und SPDX (ISO/IEC 5962).

MHC-03 — Phishing-resistente Multi-Faktor-Authentisierung

Auf einen Blick

- Operativer Nutzen: Macht abgefangene oder erbeutete Anmeldedaten auf gefälschten Seiten wertlos und entzieht damit KI-gestütztem, täuschend echtem Phishing die Grundlage.
- Betroffene Richtlinie: Richtlinie zur Zugriffssteuerung und Authentisierung.

- Abhängigkeiten: keine Vorbedingung; wirkt gut zusammen mit MHC-04 (privilegierter Zugriff).
- Aufwandstreiber: hängt von der Zahl der Nutzer, Dienste und Altsysteme sowie von der vorhandenen Identitätsplattform ab. Der konkrete Aufwand richtet sich nach den Ressourcen der Organisation.
- Primäre Kennzahl: Anteil der Anmeldungen über phishing-resistente Verfahren, insbesondere bei privilegierten und extern erreichbaren Zugängen.
- Berührte Standards (ohne Erfüllungsanspruch): ISO/IEC 27002 A.5.17/A.8.5; NIST SP 800-63B Revision 4; FIDO2/WebAuthn; C5:2026 IAM-08; NIS2-DVO Nr. 11.2.

1. Control-Statement

Privilegierte Zugänge, extern erreichbare Dienste und Zugänge zu besonders schützenswerten Systemen werden mit phishing-resistenten Verfahren abgesichert (FIDO2/WebAuthn, Passkeys oder Hardware-Token). SMS- und einfache Bestätigungsverfahren werden für neue Zugänge ausgeschlossen.

2. Zweck und Bedrohungsbezug

Phishing ist durch KI in Qualität und Menge stark gestiegen: gefälschte Anmeldeseiten, täuschend echte E-Mails und in Echtzeit weitergereichte Codes umgehen klassische Zwei-Faktor-Verfahren wie SMS oder App-Bestätigung. Phishing-resistente Verfahren binden den Anmeldenachweis kryptografisch an die echte Adresse des Dienstes; auf einer gefälschten Seite ist der Nachweis daher wertlos. Schutzziel ist, dass abgefangene oder erbeutete Anmeldedaten dem Angreifer nichts nützen.

3. Organisatorische Umsetzung (CISO-Perspektive)

- Richtlinie: Die Authentisierungsrichtlinie schreibt phishing-resistente Verfahren für privilegierte, extern erreichbare und besonders schützenswerte Zugänge vor; SMS- und einfache Push-Verfahren sind für neue Zugänge ausgeschlossen, Altverfahren werden mit Übergangsplan ersetzt.
- Verantwortlichkeiten: Die Umstellung wird im Statement of Applicability einer Stelle zugewiesen (Vorlage MRIS Anhang H).
- Risikoanalyse und SoA: Der Soll-Reifegrad wird aus der Risikobewertung abgeleitet; privilegierte und nach außen erreichbare Zugänge zuerst.
- Prozesse: Ausgabe und Rücknahme der Token bzw. Passkeys als geregelter Prozess; Ersatzverfahren für Verlust, ohne das Schutzniveau zu unterlaufen; Stilllegung der Altverfahren nach dem Übergang.
- Schulung: Nutzer werden bei Einrichtung und Nutzung der neuen Verfahren begleitet, um Akzeptanz und reibungslosen Umstieg zu sichern.

4. Technische Umsetzung (für IT-Fachleute)

- Phishing-resistente Verfahren: FIDO2/WebAuthn mit Passkeys oder Hardware-Sicherheitsschlüsseln; die kryptografische Bindung an die Domäne des Dienstes (Origin-Bindung) verhindert die Weitergabe an gefälschte Seiten.
- Schutzniveaus nach NIST SP 800-63B Revision 4: Auf Stufe AAL2 sind auch synchronisierbare Passkeys (über mehrere Geräte) zulässig; für die höchste Stufe AAL3 ist ein gerätegebundener, nicht exportierbarer Schlüssel erforderlich (Hardware-Token oder Smartcard), synchronisierbare Passkeys sind dort nicht zulässig.
- Abschaltung schwacher Verfahren: SMS-Codes und einfache App-Bestätigungen ohne Nummernabgleich werden für neue Zugänge nicht mehr zugelassen; bestehende Verfahren werden ersetzt.

- Einbindung: Durchsetzung über die zentrale Identitätsplattform und bedingten Zugriff; Anwendungen werden über die Identitätsplattform angebunden statt über eigene Anmeldemasken.
- Vertiefung: Anforderungen je Schutzniveau (AAL2/AAL3) in NIST SP 800-63B Revision 4; Verfahrensdetails in den FIDO2/WebAuthn-Spezifikationen (FIDO Alliance, W3C WebAuthn).
- Wirksamkeitstest: Ein kontrollierter Phishing-Versuch mit nachgebauter Anmeldeseite wird durchgeführt; der phishing-resistente Anmeldenachweis darf sich auf der gefälschten Seite nicht verwenden lassen (die Anmeldung schlägt fehl).

5. Umsetzungsbeispiele

Beispiel A — Entwicklungs-/Plattform-Kontext. Ein Unternehmen sichert die Zugänge seiner Administratoren und Entwickler bisher mit Passwort und einem per App bestätigten zweiten Faktor. Ein gut gemachter Phishing-Angriff fängt sowohl Passwort als auch Bestätigung in Echtzeit ab. Das Unternehmen rüstet zunächst alle privilegierten und von außen erreichbaren Zugänge auf Hardware-Sicherheitsschlüssel um, deren Nachweis kryptografisch an die echte Adresse des Dienstes gebunden ist. Auf einer gefälschten Anmeldeseite ist dieser Nachweis nutzlos. In einer weiteren Ausbaustufe wird die Anmeldung organisationsweit passwortlos, und die alten SMS- und App-Verfahren werden abgeschaltet.

Beispiel B — Allgemeine Abteilung. Eine Organisation ohne eigene Entwicklung nutzt Microsoft 365 und einige weitere Cloud-Dienste. Die Beschäftigten melden sich bisher mit Passwort und SMS-Code an — anfällig für Phishing. Die Organisation aktiviert in ihrer Identitätsplattform die Anmeldung per Passkey und gibt für besonders schützenswerte Rollen Hardware-Sicherheitsschlüssel aus. Die Anmeldung erfolgt künftig ohne Passwort und ohne SMS; eine zentrale Zugriffsregel verlangt für sensible Anwendungen ausdrücklich ein phishing-resistentes Verfahren. So lässt sich der Schutz auch ohne eigene Entwicklung über die Einstellungen der genutzten Plattform durchsetzen.

6. Reifegrad-Pfad (kumulativ)

- Initial: FIDO2/Passkeys für privilegierte Zugänge.
- Defined: phishing-resistente Verfahren verpflichtend für alle extern erreichbaren Dienste.
- Managed: passwortlose Anmeldung organisationsweit; SMS-Verfahren abgeschaltet.

7. Messung und Audit-Nachweis

- Kennzahl: Anteil der Anmeldungen über phishing-resistente Verfahren, getrennt nach privilegierten, extern erreichbaren und übrigen Zugängen (Zielwert: vollständig bei privilegierten und extern erreichbaren Zugängen).
- Nachweis: Konfiguration der Identitätsplattform, Liste der zugelassenen Verfahren je Zugangsklasse, Nachweis der Abschaltung von SMS-Verfahren.
- Prüflogik nach ISO/IEC 27005: dokumentiert? — Authentisierungsrichtlinie und Plattformkonfiguration; verantwortlich? — Identitäts-/IT-Leitung; Häufigkeit? — fortlaufend; Toleranz? — keine privilegierten oder extern erreichbaren Zugänge mit ausschließlich SMS- oder einfacher Push-Bestätigung.

8. Typische Fehler

- Phishing-resistente Verfahren werden eingeführt, aber schwächere Verfahren bleiben als Ausweichweg aktiv — der Angreifer nutzt den schwächsten Pfad.
- Für die höchste Schutzstufe werden synchronisierbare Passkeys eingesetzt, obwohl dort ein gerätegebundener, nicht exportierbarer Schlüssel erforderlich ist.

- Das Ersatzverfahren bei Verlust (etwa Rücksetzung per Anruf) unterläuft das Schutzniveau.
- Nur einzelne Anwendungen werden umgestellt, während privilegierte Zugänge weiterhin schwächer abgesichert bleiben.

9. Abgrenzung, Restrisiko und Verweise

- Abgrenzung: betrifft den Anmeldenachweis der Nutzer; die Identität von Diensten und Workloads deckt MHC-04 ab.
- Restrisiko: Angriffe können sich auf die Wiederherstellungs- und Ausweichverfahren verlagern; auch Sitzungsübernahmen nach erfolgter Anmeldung bleiben ein eigenes Thema.
- Wirksamkeitsgrenze: phishing-resistent bedeutet nicht phishing-sicher. FIDO2/Passkeys entwerten klassisches Credential-Phishing und MFA-Müdigkeit, nicht aber Sitzungs- oder Token-Diebstahl per Malware, Adversary-in-the-Middle nach der Anmeldung oder Social Engineering am Helpdesk.
- ISO/IEC 27002:2022: A.5.17 Authentisierungsinformationen; A.8.5 Sichere Authentisierung; A.6.7 Remote-Arbeit; A.8.1 Endpunktgeräte des Benutzers; A.8.4 Zugriff auf den Quellcode; A.8.23 Webfilterung.
- Framework: NIST SP 800-63B Revision 4 (Schutzniveaus AAL2/AAL3); FIDO2/WebAuthn (FIDO Alliance, W3C); C5:2026 IAM-08; NIS2-DVO Nr. 11.2.

MHC-04 — Workload-Identität und Zero-Trust-Netzwerkarchitektur

Auf einen Blick

- Operativer Nutzen: Verhindert die seitliche Ausbreitung eines Angreifers nach einer Erstkompromittierung; ein übernommener Dienst kann sich nicht mehr zu Nachbardiensten weiterbewegen.
- Betroffene Richtlinie: Zugriffssteuerungs- und Netzwerksicherheitsrichtlinie.
- Abhängigkeiten: bildet die Identitätsgrundlage für MHC-13; selbst ohne Vorbedingung.
- Aufwandstreiber: Die Umstellung bestehender Dienste bindet mehr Ressourcen als ein Neubau; der konkrete Aufwand hängt von den verfügbaren Ressourcen der Organisation ab.
- Primäre Kennzahl: Anteil der produktiven Dienst-zu-Dienst-Verbindungen mit Workload-Identität und mTLS.
- Berührte Standards (ohne Erfüllungsanspruch): u. a. ISO/IEC 27002 A.8.20/A.8.21/A.8.22, NIST SP 800-207 und 800-207A, NIS2-DVO Nr. 6.7/8/11.

1. Control-Statement

Jede Workload (Dienst, Prozess, Container, AI-Agent) erhält eine kryptografisch verifizierbare, kurzlebige und nicht wiederverwendbare Identität. Die Kommunikation zwischen Diensten wird anhand dieser Identität autorisiert, nicht anhand der Netzwerklage (IP-Adresse, Subnetz, Perimeter).

2. Zweck und Bedrohungsbezug

Mit AI beschleunigte Angreifer gelangen von der Erstkompromittierung zur seitlichen Ausbreitung (Lateral Movement) innerhalb von Minuten. Klassische Perimeter-, VPN- und Jump-Host-Modelle vertrauen einem Dienst allein wegen seiner Netzwerklage; ein übernommener Host erhält dadurch faktisch Zugriff auf die Nachbardienste. Eine eigene Identität je Workload entzieht diesem Muster

die Grundlage: ohne gültige, kurzlebige Identität kommt keine Verbindung zustande. Ziel ist, die klassischen Pfade für die seitliche Ausbreitung strukturell zu schließen. Der Schutz beruht auf einer kryptografischen Barriere, nicht darauf, dass ein Angriff nur verzögert wird.

3. Organisatorische Umsetzung (CISO-Perspektive)

- Richtlinie: Die Richtlinie zur Zugriffssteuerung und Netzwerksicherheit wird um den Grundsatz ergänzt, dass produktive Dienste sich über eine eigene, technisch verwaltete Identität ausweisen; auf reine Netzwerkadressen gestütztes Vertrauen ist nur befristet und dokumentiert zulässig.
- Verantwortlichkeiten: Die Maßnahme wird im Statement of Applicability einer benannten Stelle zugewiesen (Vorlage MRIS Anhang H: IT-Leitung verantwortlich für den Aufbau, Entwicklung verantwortlich für die Umstellung der Dienste, CISO rechenschaftspflichtig).
- Risikoanalyse und SoA: Der Soll-Reifegrad (Initial/Defined/Managed) wird aus der Risikobewertung abgeleitet; Dienste mit Zugriff auf schützenswerte Daten zuerst. Behandeltes Risiko: seitliche Ausbreitung nach einer Erstkompromittierung (Anbindung an die Risikobewertungs-Brücke, MRIS Anhang F).
- Prozesse: Es wird ein dokumentierter Prozess für Vergabe, regelmäßige Erneuerung und Entzug der Dienst-Identitäten benötigt; die Zeit bis zum Entzug einer Identität wird als Kennzahl geführt.
- Schulung: Entwicklungsteams werden auf die geänderte Vorgabe verpflichtet (keine statischen Zugangsschlüssel im Code) und eingewiesen.
- AI-Agenten: Es wird festgelegt, dass AI-Agenten unter einer eigenen, technisch verwalteten Identität arbeiten und nicht unter persönlichen Benutzerkonten; die Detailregelung erfolgt in MHC-13.

4. Technische Umsetzung (für IT-Fachleute)

- Workload-Identität mit SPIFFE/SPIRE: Jede Workload erhält eine SPIFFE-ID, ausgestellt als kurzlebiges X.509-SVID oder JWT-SVID durch eine SPIRE-Control-Plane (Server) mit einem SPIRE-Agent je Knoten. Die Workload wird vor der Ausgabe über Plattform-Selektoren attestiert (z. B. Kubernetes-Service-Account, Node-Attestation). SPIFFE/SPIRE sind ein CNCF-graduated (produktionsreifes) Projekt.
- mTLS: Beide Endpunkte authentisieren sich gegenseitig über ihre SVIDs auf Basis von TLS 1.3. Kurze Zertifikatslaufzeiten (Minuten bis Stunden) mit automatischer Rotation; langlebige gemeinsame Geheimnisse (Shared Secrets) entfallen.
- Service Mesh (z. B. Istio, Linkerd): Ein Sidecar-Proxy je Dienst übernimmt mTLS, Identitätsprüfung und Policy-Durchsetzung transparent zum Anwendungscode. Autorisiert wird über Identitäts-Policies (welche Dienst-Identität darf welchen Dienst aufrufen) statt über IP-Allowlists.
- Segmentierung: Identitätsbasierte Autorisierung mit Default-Deny zwischen Diensten ergänzt oder ersetzt die netzbasierte Mikrosegmentierung.
- Externer und privilegierter Zugriff: ZTNA bzw. Identity-Aware Proxy statt klassischem VPN; der Zugriff wird an Nutzer-Identität, Geräte-Identität und Kontext gebunden.
- Migration: Der Permissive-Mode (Klartext und mTLS parallel) dient nur als kurzer Messzeitraum; danach wird der Strict-Mode erzwungen. Reihenfolge: die kritischsten Dienst-zu-Dienst-Pfade zuerst.
- AI-Agenten: zeitlich begrenzte, auf die nötigen Fähigkeiten begrenzte (capability-scoped) Identitäten je Werkzeugaufruf statt Entwickler-Credentials.
- Vertiefung: Architekturmodell (Identity-Tier vs. Network-Tier, Ingress-/Egress-Gateways, Multi-Cloud) in NIST SP 800-207A, Abschn. 3–4; Zero-Trust-Grundprinzipien in NIST SP

800-207, Abschn. 2; SPIFFE-Spezifikation (SPIFFE-ID, X.509-SVID, JWT-SVID) unter spiffe.io.
- Wirksamkeitstest: Von einem Dienst oder Host ohne gültige Identität wird eine Verbindung zu einem geschützten Dienst versucht — sie muss abgewiesen werden. Ergänzend: Ein abgelaufener Ausweis darf keine Verbindung mehr ermöglichen.

5. Umsetzungsbeispiele

Beispiel A — Entwicklungs-/Plattform-Kontext. Ein mittelständischer Software-Anbieter betreibt seine Anwendung aus vielen einzelnen, miteinander vernetzten Diensten. Bisher vertrauten sich diese Dienste gegenseitig allein deshalb, weil sie im selben internen Netz liefen; übernahm ein Angreifer einen Dienst, gelangte er von dort ungehindert an alle übrigen. Das Unternehmen ändert diese Grundlage: Jeder Dienst erhält einen eigenen, ständig wechselnden technischen Ausweis, und ohne gültigen Ausweis nimmt kein anderer Dienst eine Verbindung an. Begonnen wird mit den wenigen Diensten, die Kundendaten verarbeiten; später wird die Umstellung auf alle ausgeweitet. Das Ergebnis: Ein übernommener Dienst steht für sich allein und kann sich nicht mehr seitwärts zu den Nachbardiensten weiterbewegen.

Beispiel B — Allgemeine Abteilung. Eine Organisation ohne eigene Softwareentwicklung betreibt einige interne Anwendungen, etwa einen Dateiserver, und gewährt den Zugriff bisher über das interne Netz und ein VPN: Wer im Netz ist, gilt als vertrauenswürdig. Das birgt dasselbe Grundproblem — ein übernommenes Gerät im Netz erhält weitreichenden Zugriff. Die Organisation bindet den Zugriff auf interne Anwendungen künftig an die geprüfte Identität von Nutzer und Gerät, nicht mehr an die bloße Zugehörigkeit zum Netz; ein vorgeschalteter Zugangsdienst prüft bei jedem Zugriff Anmeldung, Gerät und Situation. Begonnen wird mit den schützenswertesten Anwendungen und den Administrationszugängen. *Anwendbarkeitsnotiz:* Eine Organisation, die ausschließlich fertige Cloud-Dienste nutzt und keine eigene Infrastruktur betreibt, ist von MHC-04 kaum betroffen; im Statement of Applicability wird das Control als nicht anwendbar geführt. Die Absicherung verantwortet hier der Anbieter und wird vertraglich über das Lieferantenmanagement eingefordert.

6. Reifegrad-Pfad (kumulativ; kritische Pfade zuerst, kein Big-Bang)

- Initial: SPIFFE/SPIRE für privilegierte Workloads; mTLS auf den 3–5 kritischsten Dienst-zu-Dienst-Pfaden.
- Defined: Workload-Identität für alle produktiven Microservices; ZTNA für privilegierten Remote-Zugriff.
- Managed: flächendeckend inklusive Dev/Test; identitätsbasierte Mikrosegmentierung; AI-Agenten mit capability-scoped Identitäten.

7. Messung und Audit-Nachweis

- Kennzahlen: Anteil der produktiven Dienst-zu-Dienst-Verbindungen mit mTLS/Workload-Identität (Zielwert Defined: 100 % produktiv); Zeit bis zum Entzug einer Workload-Identität.
- Nachweis: SPIRE-Registry (Inventar der ausgestellten Identitäten), Service-Mesh-Policy-Konfiguration, Protokolle der mTLS-Durchsetzung.
- Prüflogik nach ISO/IEC 27005: dokumentiert? — über die SPIRE-Registry; verantwortlich? — Plattform-Team; Häufigkeit? — fortlaufend (Erneuerung im Minuten- bis Stundentakt); Toleranz? — keine produktiven, im Klartext gespeicherten gemeinsamen Geheimnisse auf kritischen Pfaden.

8. Typische Fehler

- mTLS bleibt im Permissive-Mode (Klartext wird weiter akzeptiert) — keine Schutzwirkung.
- Identitäten mit zu langer Gültigkeit (Tage statt Minuten) — der Vorteil der kurzen Laufzeit entfällt.
- AI-Agenten arbeiten weiter unter persönlichen Entwickler-Konten — Nachvollziehbarkeit und Begrenzung der Rechte fehlen.
- Freigaben auf IP-Basis als dauerhafter Parallelweg statt als befristete, dokumentierte Ausnahme.

9. Abgrenzung, Restrisiko und Verweise

- Abgrenzung: schützt die Ebene Dienst-zu-Dienst; die Authentisierung der Endnutzer deckt MHC-03 ab; die Governance der AI-Agenten vertieft MHC-13.
- Restrisiko: eine übernommene Ausstellungs-Infrastruktur (SPIRE) oder ein legitim ausgewiesener, aber übernommener Dienst bleiben wirksam; eine Identität ersetzt nicht die Minimierung der Rechte (Least Privilege bleibt erforderlich).
- Wirksamkeitsgrenze: Zero Trust ist ein Architektur-Weg, kein Produkt. Eine nur teilweise umgesetzte Segmentierung oder Workload-Identität kann falsche Sicherheit erzeugen; Wirksamkeit entsteht erst, wenn Authentisierung, Autorisierung und Least Privilege durchgängig und prüfbar greifen.
- ISO/IEC 27002:2022: A.5.15 Zugangssteuerung; A.5.16 Identitätsmanagement; A.6.7 Remote-Arbeit; A.8.2 Privilegierte Zugangsrechte; A.8.20 Netzwerksicherheit; A.8.21 Sicherheit von Netzwerkdiensten; A.8.22 Trennung von Netzwerken.
- Framework: NIST SP 800-207 (Zero-Trust-Prinzipien); NIST SP 800-207A (Cloud-native Zero Trust, Workload-Identität, Service Mesh); SPIFFE/SPIRE (CNCF).

MHC-05 — Verhaltensbasierte Detection und Kill-Chain-Korrelation

Auf einen Blick

- Operativer Nutzen: Erkennt Angreifer an ihrem Verhalten und an der Kette zusammenhängender Schritte, statt nur an bekannten Signaturen — und fängt so auch getarnte, mit legitimen Mitteln geführte Angriffe.
- Betroffene Richtlinie: Richtlinie zur Sicherheitsüberwachung (Logging, Monitoring, Detection).
- Abhängigkeiten: setzt zentrale, manipulationssichere Protokollierung voraus (A.8.15); liefert die Signale für die Automatisierung in MHC-11.
- Aufwandstreiber: hängt von Umfang der Protokollquellen, der vorhandenen SIEM-/Detection-Plattform und der verfügbaren Analyse-Kapazität ab. Der konkrete Aufwand richtet sich nach den Ressourcen der Organisation.
- Primäre Kennzahl: Abdeckung der relevanten Angriffstechniken nach MITRE ATT&CK (Anteil der wichtigsten Techniken mit wirksamer Erkennung).
- Berührte Standards (ohne Erfüllungsanspruch): ISO/IEC 27002 A.8.16/A.5.7; MITRE ATT&CK; C5:2026 OPS-13; NIS2-DVO Nr. 3.2; DORA Art. 10.

1. Control-Statement

Angriffe werden anhand von Verhaltensauffälligkeiten und der Verknüpfung zusammenhängender Ereignisse erkannt, nicht allein anhand einzelner bekannter Signaturen. Die Erkennung orientiert sich an einem anerkannten Katalog von Angriffstechniken (MITRE ATT&CK) mit messbarer

Abdeckung; die gezielte Suche nach noch unentdeckten Angriffen (Threat-Hunting) ist eine feste Funktion.

2. Zweck und Bedrohungsbezug

KI-gestützte Angreifer gehen unauffällig vor: Sie nutzen bordeigene Systemwerkzeuge statt mitgebrachter Schadsoftware und tarnen ihre Steuerkanäle, sodass jeder einzelne Schritt für sich harmlos wirkt. Klassische, signaturbasierte Erkennung schlägt dabei nicht an. Außerdem zerlegen agentische Angreifer ihr Vorgehen in viele kleine, je für sich unverdächtige Aktionen — die Gefahr zeigt sich erst im Zusammenhang. Schutzziel ist, Angriffe an ihrem Verhalten und an der Abfolge zusammenhängender Schritte zu erkennen, auch wenn keine bekannte Signatur vorliegt.

3. Organisatorische Umsetzung (CISO-Perspektive)

- Richtlinie: Die Überwachungsrichtlinie legt fest, dass die Erkennung verhaltens- und technikbasiert erfolgt (Bezug auf MITRE ATT&CK), mit messbarer Abdeckung und regelmäßiger Überprüfung.
- Verantwortlichkeiten: Detection-Betrieb und Threat-Hunting werden im Statement of Applicability zugewiesen (Vorlage MRIS Anhang H); Threat-Hunting ist als eigene Funktion mit Mindestkapazität verankert.
- Risikoanalyse und SoA: Die wichtigsten zu erkennenden Techniken werden aus Bedrohungslage und Risikobewertung abgeleitet. Behandeltes Risiko: getarnte, mit legitimen Mitteln geführte Angriffe, die unter der Schwelle einzelner Alarme bleiben.
- Prozesse: dokumentierte Threat-Hunting-Methodik (z. B. PEAK oder TaHiTI) mit regelmäßigen, hypothesengestützten Durchläufen; regelmäßige Überprüfung und Erweiterung der Erkennungsabdeckung; geregelte Übergabe erkannter Vorfälle in die Reaktion.
- Kapazität: ausreichende Analyse-Kapazität für Hunting und Auswertung; Ausrichtung des Sicherheitsbetriebs auf Analyse statt reiner Alarmsichtung.

4. Technische Umsetzung (für IT-Fachleute)

- Verhaltensbasierte Erkennung: Auswertung von Nutzer- und Systemverhalten gegen eine erlernte Normallinie (UEBA), mit Alarmierung bei Abweichungen.
- Technik-Abdeckung nach MITRE ATT&CK: Erkennungsregeln werden an den Techniken des ATT&CK-Katalogs ausgerichtet (laufend aktualisiert, aktuell v19) statt an isolierten Einzelsignaturen; die Abdeckung wird gemessen (z. B. mit DeTT&CT) und durch kontrollierte Angriffstests bestätigt (z. B. Atomic Red Team). Richtwert: die wichtigsten verbreiteten Techniken zuverlässig abgedeckt.
- Kill-Chain-Korrelation: einzelne Ereignisse werden über Zeit und Systeme hinweg zu einer Angriffskette verknüpft, sodass eine Folge je für sich unauffälliger Schritte als Ganzes auffällt.
- Erkennungsschwerpunkte: Nutzung bordeigener Werkzeuge (z. B. auffällige Skript- oder Aufgabenplanungs-Aktivität, ungewöhnliche Systemprozesse) und getarnte Steuerkanäle (z. B. ungewöhnlicher DNS- oder verschlüsselter Datenverkehr mit langen Ruhephasen).
- Robustheit der eigenen Detection-KI: setzt die Erkennung selbst auf maschinelles Lernen, werden die Modelle gegen Täuschung (Adversarial Examples) und Datenvergiftung geprüft.
- Vertiefung: Technikkatalog und Erkennungshinweise in MITRE ATT&CK (Enterprise); Hintergrund zu Erkennungssystemen in NIST SP 800-94 (Fassung 2007; eine überarbeitete Fassung liegt nicht vor).
- Wirksamkeitstest: Ein mehrstufiger, kontrollierter Angriff aus je für sich unauffälligen Schritten (z. B. über Atomic Red Team) wird durchgeführt; die Kette muss als

zusammenhängender Vorfall erkannt werden, nicht nur als einzelne, unverbundene Ereignisse.

5. Umsetzungsbeispiele

Beispiel A — Entwicklungs-/Plattform-Kontext. Ein Unternehmen mit eigenem Sicherheitsbetrieb stellt fest, dass seine bisherige Erkennung nur auf bekannte Schadsoftware anspringt. Ein Angreifer, der ausschließlich bordeigene Systemwerkzeuge nutzt, bleibt unentdeckt. Das Unternehmen richtet seine Erkennungsregeln deshalb an einem anerkannten Katalog von Angriffstechniken aus und misst, welcher Anteil der wichtigsten Techniken tatsächlich abgedeckt ist. Verdächtige Schrittfolgen werden über Zeit und Systeme hinweg zu einer Kette verknüpft. Zusätzlich sucht ein kleines Team regelmäßig und gezielt nach Spuren bisher unentdeckter Angriffe. So fällt nun auch ein Angreifer auf, der sich mit legitimen Mitteln leise bewegt.

Beispiel B — Allgemeine Abteilung. Eine Organisation ohne eigenen 24/7-Betrieb kann eine solche Erkennung nicht selbst aufbauen. Sie bezieht die Sicherheitsüberwachung deshalb als Dienstleistung (Managed Detection & Response) und achtet bei der Auswahl darauf, dass der Anbieter verhaltens- und technikbasiert erkennt, zusammenhängende Angriffsketten auswertet und vertraglich eine feste Reaktionszeit zusichert. Intern benennt sie eine Ansprechperson, die die gemeldeten Vorfälle aufnimmt und die Behebung steuert. So erreicht sie ein vergleichbares Schutzniveau über einen Dienstleister.

6. Reifegrad-Pfad (kumulativ)

- Initial: MITRE ATT&CK als Erkennungsrahmen eingeführt; geringe Abdeckung.
- Defined: die wichtigsten verbreiteten Techniken zuverlässig abgedeckt; dokumentiertes Threat-Hunting.
- Managed: gegen Täuschung robuste, lernende Erkennung; fortlaufende Überprüfung und Bestätigung der Abdeckung.

7. Messung und Audit-Nachweis

- Kennzahl: Abdeckung der wichtigsten Angriffstechniken nach MITRE ATT&CK (gemessen und durch Angriffstests bestätigt); ergänzend Zahl und Ergebnis der durchgeführten Threat-Hunts.
- Nachweis: Abdeckungsübersicht (z. B. DeTT&CT), Ergebnisse der Angriffstests, dokumentierte Hunting-Durchläufe mit Technik-Zuordnung.
- Prüflogik nach ISO/IEC 27005: dokumentiert? — Überwachungsrichtlinie, Abdeckungsübersicht, Hunting-Protokolle; verantwortlich? — Detection-/SOC-Leitung; Häufigkeit? — fortlaufender Betrieb, regelmäßige Hunts und Abdeckungsprüfung; Toleranz? — keine dauerhaft unbeobachteten wichtigen Techniken.

8. Typische Fehler

- Die Erkennung bleibt signaturbasiert; Angriffe mit bordeigenen Werkzeugen lösen keinen Alarm aus.
- Die ATT&CK-Abdeckung wird behauptet, aber nie durch kontrollierte Angriffstests überprüft — die Erkennung greift im Ernstfall nicht.
- Einzelne Alarme werden abgearbeitet, aber nie zu Ketten verknüpft; die Zusammenhänge bleiben unsichtbar.
- Threat-Hunting wird als Nebenaufgabe behandelt und mangels Kapazität faktisch nicht durchgeführt.

9. Abgrenzung, Restrisiko und Verweise

- Abgrenzung: betrifft das Erkennen von Angriffen; die anschließende automatisierte Reaktion behandelt MHC-11, das automatisierte Testen der eigenen Abwehr MHC-09.
- Restrisiko: Erkennung ist nicht Verhinderung — sehr langsame, stark getarnte oder neuartige Angriffe können unentdeckt bleiben; setzt die Erkennung auf maschinelles Lernen, bleibt die gezielte Täuschung der Modelle ein Restrisiko.
- ISO/IEC 27002:2022: A.5.3 Aufgabentrennung; A.5.7 Informationen über die Bedrohungslage; A.5.14 Informationsübermittlung; A.5.15 Zugangssteuerung; A.8.1 Endpunktgeräte des Benutzers; A.8.3 Informationszugangsbeschränkung; A.8.4 Zugriff auf den Quellcode; A.8.7 Schutz gegen Schadsoftware; A.8.12 Verhinderung von Datenlecks; A.8.16 Überwachung von Aktivitäten.
- Framework: MITRE ATT&CK (Enterprise, laufend aktualisiert); C5:2026 OPS-13; NIS2-DVO Nr. 3.2; DORA Art. 10; NIST SP 800-94 (Fassung 2007); Methodik-Beispiele PEAK und TaHiTI, Abdeckungsmessung mit DeTT&CT, Angriffstests mit Atomic Red Team.

MHC-06 — Container-Sicherheit und Confidential Computing

Auf einen Blick

- Operativer Nutzen: Stellt sicher, dass nur unveränderte, geprüfte Container-Images laufen, und schützt besonders sensible Daten sogar während der Verarbeitung — auch gegenüber einem Angreifer mit Zugriff auf die darunterliegende Infrastruktur.
- Betroffene Richtlinie: Richtlinie zur sicheren Entwicklung und zum sicheren Betrieb (Container- und Cloud-Betrieb).
- Abhängigkeiten: keine Vorbedingung; setzt eine Container-/Cloud-Plattform voraus; wirkt zusammen mit MHC-02 (Herkunft der Images) und MHC-04 (Identitäten).
- Aufwandstreiber: hängt von der Zahl der Container-Workloads und davon ab, ob Confidential Computing in der genutzten Plattform verfügbar ist. Der konkrete Aufwand richtet sich nach den Ressourcen der Organisation.
- Primäre Kennzahl: Anteil der produktiv betriebenen Container aus signierten, vor dem Start geprüften Images; ergänzend Abdeckung von Confidential Computing bei hochsensiblen Workloads.
- Berührte Standards (ohne Erfüllungsanspruch): ISO/IEC 27002 A.8.27/A.8.31; C5:2026 OPS-32 bis OPS-35; NIST SP 800-190; Confidential Computing Consortium.

1. Control-Statement

Container laufen nur aus signierten, vor dem Start geprüften Images aus kontrollierten Quellen, mit Durchsetzung über die Plattform. Für Workloads mit besonders hohem Schutzbedarf werden geschützte Ausführungsumgebungen (Confidential Computing) mit Echtheitsnachweis (Remote-Attestation) eingesetzt.

MHC-06 umfasst zwei unterschiedlich reife und unterschiedlich breit anwendbare Schutzebenen: Container-Sicherheit als breite Basisanforderung für containerisierte Workloads und Confidential Computing als weiterführende Härtung für besonders schutzbedürftige Workloads, Mandantentrennung, regulierte Datenverarbeitung oder Infrastruktur mit erhöhtem Vertrauensrisiko. Beide werden hier gemeinsam behandelt, sollten in der Umsetzung aber getrennt geplant, bewertet und nachgewiesen werden.

2. Zweck und Bedrohungsbezug

Container werden oft aus öffentlich verfügbaren Basis-Images zusammengesetzt. Ist ein solches Image manipuliert, läuft der Schadcode unbemerkt mit. KI erleichtert es Angreifern, manipulierte

Images und passende Schwachstellen zu finden. Zudem sind in geteilten Cloud-Umgebungen Daten zwar bei Übertragung und Speicherung verschlüsselt, während der Verarbeitung im Arbeitsspeicher aber grundsätzlich einsehbar — ein Angreifer mit Zugriff auf die Infrastruktur könnte sie dort abgreifen. Schutzziel ist, nur unveränderte, geprüfte Container auszuführen und besonders sensible Daten auch während der Verarbeitung zu schützen.

3. Organisatorische Umsetzung (CISO-Perspektive)

- Richtlinie: Die Betriebsrichtlinie schreibt signierte, geprüfte Images aus kontrollierten Quellen vor sowie Confidential Computing für klar benannte, besonders schützenswerte Workloads.
- Verantwortlichkeiten: Im Statement of Applicability zugewiesen (Vorlage MRIS Anhang H); Plattform-/Betriebsteam und Sicherheit wirken zusammen.
- Risikoanalyse und SoA: Welche Workloads Confidential Computing benötigen, wird aus dem Schutzbedarf abgeleitet (nicht alles braucht es). Behandeltes Risiko: manipulierte Images sowie Zugriff auf Daten in der Verarbeitung (Anbindung an die Risikobewertungs-Brücke, MRIS Anhang F).
- Prozesse: kontrollierte Image-Quellen und Freigaben; geregelter Umgang mit Schwachstellenfunden vor dem Deployment; bei Bezug aus der Cloud Nachweis der Provider-Fähigkeiten (Confidential Computing, Attestation).

4. Technische Umsetzung (für IT-Fachleute)

- Signierte Images: Basis- und eigene Images werden signiert (z. B. mit Sigstore/cosign) und aus kontrollierten Registries bezogen; die Signatur wird vor dem Start geprüft.
- Durchsetzung beim Deployment: ein Admission-Controller der Plattform (z. B. OPA Gatekeeper oder Kyverno in Kubernetes) lässt nur signierte, geprüfte Images zu und blockiert den Rest.
- Image-Prüfung: automatisierte Schwachstellen-Scans der Images vor dem Deployment; regelmäßiges erneutes Scannen gelagerter Images, wenn neue Schwachstellen bekannt werden.
- Confidential Computing: für hochsensible Workloads geschützte Ausführungsumgebungen (TEE/Secure Enclaves, z. B. Intel TDX, AMD SEV-SNP, ARM CCA), die Daten auch im Arbeitsspeicher verschlüsseln; vor der Nutzung wird per Remote-Attestation nachgewiesen, dass die Umgebung echt und unverändert ist.
- Vertiefung: Container-Risiken und Schutzmaßnahmen in NIST SP 800-190 (Fassung 2017, weiterhin maßgebliche Referenz); herstellerübergreifende Grundlagen beim Confidential Computing Consortium; Anforderungen in C5:2026 OPS-32 bis OPS-35.
- Wirksamkeitstest: Es wird versucht, ein nicht signiertes oder nicht geprüftes Image in der produktiven Umgebung zu starten — der Start muss durch die Plattform verhindert werden.

5. Umsetzungsbeispiele

Beispiel A — Entwicklungs-/Plattform-Kontext. Ein Anbieter betreibt seine Software in Containern, die zu großen Teilen aus öffentlichen Basis-Images bestehen. Bisher prüft niemand, ob diese Images unverändert sind. Der Anbieter führt deshalb durchgängige Image-Signaturen ein: Nur signierte, zuvor auf Schwachstellen geprüfte Images aus den eigenen kontrollierten Registries dürfen starten, durchgesetzt über einen Admission-Controller der Plattform. Für die wenigen Dienste, die besonders sensible Kundendaten verarbeiten, nutzt er zusätzlich eine geschützte Ausführungsumgebung, in der die Daten auch im Arbeitsspeicher verschlüsselt bleiben, und weist deren Echtheit vor der Nutzung nach. Manipulierte Images werden so abgewiesen, und selbst ein Angreifer mit Infrastruktur-Zugriff kommt an die sensibelsten Daten nicht heran.

Anwendbarkeitsnotiz (statt Beispiel B). Container-Sicherheit und Confidential Computing setzen voraus, dass die Organisation Container selbst baut oder betreibt. Eine Organisation, die ausschließlich fertige SaaS-Dienste nutzt, kann diese Maßnahme nicht selbst umsetzen; für sie wird sie zur Beschaffungs- und Nachweisfrage: Beim Anbieter wird nachgehalten, ob er Images signiert und prüft, den Start nicht freigegebener Images unterbindet und für besonders sensible Daten geschützte Ausführungsumgebungen anbietet — belegt etwa durch Zertifikate oder Prüfberichte (z. B. C5).

6. Reifegrad-Pfad (kumulativ)

- Initial: Container-Images signiert; Schwachstellen-Scan vor dem Deployment.
- Defined: Durchsetzung über Admission-Controller im Betrieb; Einsatzfälle für Confidential Computing benannt und dokumentiert.
- Managed: Confidential Computing produktiv für hochsensible Workloads; Remote-Attestation automatisiert.

7. Messung und Audit-Nachweis

- Kennzahl: Anteil der produktiven Container aus signierten, vor dem Start geprüften Images (Zielwert: vollständig); zusätzlich Abdeckung von Confidential Computing bei den als hochsensibel eingestuften Workloads.
- Nachweis: Konfiguration des Admission-Controllers, Signatur- und Scan-Protokolle, Attestation-Nachweise der geschützten Umgebungen.
- Prüflogik nach ISO/IEC 27005: dokumentiert? — Betriebsrichtlinie, Plattformkonfiguration; verantwortlich? — Plattform-/Betriebsleitung; Häufigkeit? — bei jedem Deployment, Scans fortlaufend; Toleranz? — kein produktiver Container ohne gültige Signatur und Prüfung.

8. Typische Fehler

- Images werden zwar signiert, aber die Signatur wird beim Start nicht geprüft — die Signatur ist dann wirkungslos.
- Der Admission-Controller ist vorhanden, läuft aber nur im Warnmodus; nicht erlaubte Images werden gemeldet, aber nicht blockiert.
- Confidential Computing wird pauschal für alles gefordert, obwohl es nur für wenige hochsensible Workloads nötig ist — unnötiger Aufwand.
- Gelagerte Images werden nach dem ersten Scan nie erneut geprüft; neu bekannt gewordene Schwachstellen bleiben unentdeckt.

9. Abgrenzung, Restrisiko und Verweise

- Abgrenzung: betrifft die Integrität der Container und den Schutz von Daten in der Verarbeitung; die Herkunft der enthaltenen Komponenten behandelt MHC-02, die Trennung mehrerer Mandanten MHC-07.
- Restrisiko: Confidential Computing schützt die Verarbeitung, nicht jedoch Fehler in der Anwendung selbst; geschützte Ausführungsumgebungen sind nicht überall verfügbar und können die Leistung beeinflussen.
- Anwendbarkeit: Container-Sicherheit ist Basisanforderung bei containerisierten Workloads; Confidential Computing ist kontextabhängig und im Statement of Applicability häufig als N/A zu begründen. Beide Schutzebenen sind getrennt zu planen, zu bewerten und nachzuweisen.
- ISO/IEC 27002:2022 (mittelbar): A.8.27 Sichere Systemarchitektur und Entwicklungsgrundsätze; A.8.31 Trennung von Entwicklungs-, Test- und Betriebsumgebungen; A.5.23 Informationssicherheit für die Nutzung von Cloud-Diensten.

- Framework: C5:2026 OPS-34/35 (Container Management), OPS-32/33 (Confidential Computing), PSS-11; NIST SP 800-190; Confidential Computing Consortium; Image-Signatur über Sigstore/cosign.

MHC-07 — Multi-Tenancy-Isolation mit nachweisbarer Trennung

Auf einen Blick

- Operativer Nutzen: Verhindert nachweislich, dass ein Angreifer, der bei einem Mandanten (Tenant) Fuß fasst, auf die Daten anderer Mandanten übergreift — die Schadwirkung bleibt auf einen Mandanten begrenzt.
- Betroffene Richtlinie: Richtlinie zur sicheren Architektur und Mandantentrennung (für mandantenfähige Dienste).
- Abhängigkeiten: keine Vorbedingung; betrifft Organisationen, die mehrere Kunden auf gemeinsamer Plattform betreiben; nutzt Schlüsseltrennung (Bezug zu A.8.24) und Netztrennung.
- Aufwandstreiber: hängt vom Architekturmodell (gemeinsame vs. getrennte Ressourcen) und der Zahl der Mandanten ab. Der konkrete Aufwand richtet sich nach den Ressourcen der Organisation.
- Primäre Kennzahl: Ergebnis regelmäßiger, möglichst automatisierter Trennungstests (kein mandantenübergreifender Zugriff nachweisbar).
- Berührte Standards (ohne Erfüllungsanspruch): ISO/IEC 27002 A.8.22/A.5.23; C5:2026 OPS-30/31, PSS-10; CSA Cloud Controls Matrix.

1. Control-Statement

In mandantenfähigen Diensten werden Daten, Netzwerke und Rechenressourcen je Mandant getrennt, und die Trennung wird durch regelmäßige, möglichst automatisierte Tests nachgewiesen — nicht nur konzeptionell behauptet.

2. Zweck und Bedrohungsbezug

Bei mandantenfähigen Diensten teilen sich viele Kunden eine gemeinsame Plattform. Gelingt es einem Angreifer, bei einem Mandanten Fuß zu fassen, ist die zentrale Frage, ob er von dort auf die Daten anderer Mandanten übergreifen kann. KI-gestützte Angreifer suchen systematisch nach genau solchen Lücken in der Trennung. Eine bloß konzeptionelle Trennung genügt nicht; entscheidend ist, dass die Trennung tatsächlich greift und das belegt ist. Schutzziel ist, die Schadwirkung eines Angriffs nachweislich auf einen einzelnen Mandanten zu begrenzen.

3. Organisatorische Umsetzung (CISO-Perspektive)

- Richtlinie: Die Architekturrichtlinie schreibt für mandantenfähige Dienste dokumentierte Trennungskonzepte für Daten, Netzwerke und Rechenressourcen sowie deren regelmäßigen Nachweis vor.
- Verantwortlichkeiten: Im Statement of Applicability zugewiesen (Vorlage MRIS Anhang H); Architektur und Betrieb wirken zusammen.
- Risikoanalyse und SoA: Das geforderte Trennungsniveau (logisch oder physisch) wird aus dem Schutzbedarf der Kundendaten abgeleitet. Behandeltes Risiko: mandantenübergreifender Zugriff (Anbindung an die Risikobewertungs-Brücke, MRIS Anhang F).
- Prozesse: regelmäßige Trennungstests mit Protokoll; Behandlung festgestellter Schwächen; Nachweisführung gegenüber Kunden und Prüfern.

4. Technische Umsetzung (für IT-Fachleute)

- Datentrennung: mandantenspezifische Schlüssel sowie logische oder — bei höchstem Schutzbedarf — physische Trennung der Datenbestände.
- Netztrennung: getrennte Netzbereiche je Mandant (z. B. eigene virtuelle Netze/VPCs, Namespaces, regelbasierte Trennung über Software-Defined Networking).
- Trennung der Rechenressourcen: mandantenbezogene Zuteilung (Scheduling), damit Workloads verschiedener Mandanten sich nicht unkontrolliert teilen.
- Nachweis: regelmäßige, möglichst automatisierte Tests, die gezielt versuchen, aus einem Mandanten heraus auf einen anderen zuzugreifen; das Ergebnis (kein Zugriff möglich) wird dokumentiert.
- Vertiefung: Trennungsanforderungen in C5:2026 OPS-30/31 (Datentrennung) und PSS-10 (Netz); Kontrollkatalog der CSA Cloud Controls Matrix; Netztrennung als Grundsatz in A.8.22.
- Wirksamkeitstest: Aus einem Testmandanten wird gezielt versucht, auf Daten oder Ressourcen eines anderen Mandanten zuzugreifen — der Zugriff muss scheitern, und das Ergebnis wird festgehalten.

5. Umsetzungsbeispiele

Beispiel A — Entwicklungs-/Plattform-Kontext. Ein SaaS-Anbieter betreibt viele Kunden auf einer gemeinsamen Plattform. Die Trennung ist bislang nur in Architekturdokumenten beschrieben, aber nie überprüft. Der Anbieter führt mandantenspezifische Schlüssel ein, trennt die Netzbereiche je Kunde und sorgt dafür, dass sich Rechenlasten verschiedener Kunden nicht unkontrolliert teilen. Vor allem aber richtet er regelmäßige, automatisierte Tests ein, die aus einem Mandanten heraus den Zugriff auf andere zu erzwingen versuchen. Schlägt ein solcher Versuch fehl, ist die Trennung belegt; gelingt er, gibt es einen klaren Befund zum Beheben. Aus einer behaupteten wird so eine nachgewiesene Trennung.

Anwendbarkeitsnotiz (statt Beispiel B). Diese Maßnahme richtet sich an Organisationen, die selbst mandantenfähige Dienste betreiben. Eine Organisation, die solche Dienste nur nutzt, setzt die Trennung nicht selbst um; für sie wird sie zur Beschaffungs- und Nachweisfrage: Beim Anbieter wird nachgehalten, wie er die Mandantentrennung umsetzt und nachweist (etwa durch Prüfberichte oder Zertifikate wie C5) und ob die Trennung regelmäßig getestet wird.

6. Reifegrad-Pfad (kumulativ)

- Initial: logische Mandantentrennung dokumentiert.
- Defined: mandantenspezifische Schlüssel; Netz- und Ressourcentrennung umgesetzt.
- Managed: nachweisbare Trennung durch automatisierte Tests; regelmäßige Prüfungen.

7. Messung und Audit-Nachweis

- Kennzahl: Ergebnis der regelmäßigen Trennungstests (kein mandantenübergreifender Zugriff nachweisbar); Häufigkeit und Abdeckung der Tests.
- Nachweis: Trennungskonzepte, Testprotokolle, Behebungsnachweise festgestellter Schwächen.
- Prüflogik nach ISO/IEC 27005: dokumentiert? — Trennungskonzepte und Testprotokolle; verantwortlich? — Architektur-/Betriebsleitung; Häufigkeit? — regelmäßige Tests; Toleranz? — kein nachgewiesener mandantenübergreifender Zugriff.

8. Typische Fehler

- Die Trennung ist nur konzeptionell beschrieben, aber nie durch Tests belegt.

- Die Datentrennung ist umgesetzt, die Netz- oder Ressourcentrennung aber nicht — der Übergriff erfolgt über den ungeschützten Pfad.
- Tests prüfen nur den Normalfall, nicht den gezielten Umgehungsversuch aus einem Mandanten heraus.
- Bei Erweiterungen (neue Funktionen, neue Mandanten) wird die Trennung nicht erneut überprüft.

9. Abgrenzung, Restrisiko und Verweise

- Abgrenzung: betrifft die Trennung zwischen Mandanten; den Schutz der Verarbeitung selbst behandelt MHC-06, die Identität der Workloads MHC-04.
- Restrisiko: eine gemeinsame Plattform behält eine geteilte Grundschicht; Fehler in dieser gemeinsamen Schicht können trotz Trennung mehrere Mandanten betreffen.
- ISO/IEC 27002:2022 (mittelbar): A.8.22 Trennung in Netzwerken; A.5.23 Informationssicherheit für die Nutzung von Cloud-Diensten.
- Framework: C5:2026 OPS-30/31 (Datentrennung), PSS-10 (Software Defined Networking); CSA Cloud Controls Matrix; ergänzend ISO/IEC 27017 (Cloud-Dienste).

MHC-08 — Unveränderliche Backups und Recovery-Validierung

Auf einen Blick

- Operativer Nutzen: Stellt sicher, dass nach Ransomware oder Datenmanipulation eine saubere, unveränderte Kopie vorhanden ist und die Wiederherstellung nachweislich funktioniert.
- Betroffene Richtlinie: Backup- und Wiederherstellungsrichtlinie (Teil des Notfall- und Continuity-Managements).
- Abhängigkeiten: keine Vorbedingung; getrennte Zugriffswege setzen eine geordnete Rechteverwaltung voraus.
- Aufwandstreiber: hängt von Datenmenge, Wiederherstellungszielen und der vorhandenen Backup-Infrastruktur ab. Der konkrete Aufwand richtet sich nach den Ressourcen der Organisation.
- Primäre Kennzahl: Erfolgsquote der regelmäßigen Wiederherstellungstests aus unveränderlichen Backups.
- Berührte Standards (ohne Erfüllungsanspruch): ISO/IEC 27002 A.8.13/A.8.14; C5:2026 OPS-06 bis OPS-09; DORA Art. 12; NIS2-DVO Nr. 4.1; NIST SP 800-34, SP 1800-11/-25.

1. Control-Statement

Die Organisation hält mindestens eine unveränderliche oder vom Netz getrennte Sicherungskopie vor und prüft regelmäßig durch dokumentierte Wiederherstellungstests, dass sich die Daten daraus fehlerfrei zurückspielen lassen. Die Backup-Infrastruktur ist über eigene, getrennte Zugriffswege geschützt.

2. Zweck und Bedrohungsbezug

Moderne, KI-gestützte Angriffe verschlüsseln oder verfälschen Daten und suchen gezielt auch die Backups, um eine Wiederherstellung zu verhindern. Ein Angreifer mit weitreichendem Zugriff kann online erreichbare Sicherungen mitverschlüsseln oder löschen; eine unveränderliche oder vom Netz getrennte Kopie bleibt davon unberührt. Schutzziel ist, nach einem Vorfall jederzeit auf eine saubere, unveränderte Datenbasis zurückgreifen zu können — und sicher zu sein, dass die Wiederherstellung tatsächlich gelingt.

3. Organisatorische Umsetzung (CISO-Perspektive)

- Richtlinie: Die Backup- und Wiederherstellungsrichtlinie schreibt das 3-2-1-1-0-Prinzip, getrennte Zugriffswege für die Backup-Infrastruktur und regelmäßige, dokumentierte Wiederherstellungstests vor.
- Verantwortlichkeiten: Die Maßnahme wird im Statement of Applicability einer Stelle zugewiesen (Vorlage MRIS Anhang H); die Ergebnisse der Wiederherstellungstests werden berichtet.
- Risikoanalyse und SoA: Wiederherstellungsziele (wie schnell, mit welchem Datenstand) werden aus der Risiko- und Auswirkungsbetrachtung abgeleitet; besonders geschäftskritische Daten zuerst.
- Prozesse: fester Plan für regelmäßige Wiederherstellungstests mit Protokoll; getrennte Verwaltung der Zugriffsrechte auf die Backup-Infrastruktur; Aufbewahrung und Schutz der Sicherungskopien.

4. Technische Umsetzung (für IT-Fachleute)

- 3-2-1-1-0-Prinzip: drei Kopien der Daten, auf zwei verschiedenen Medientypen, mindestens eine außer Haus (offsite), mindestens eine unveränderlich oder vom Netz getrennt (immutable bzw. air-gapped), null Fehler beim regelmäßigen Wiederherstellungstest.
- Unveränderlichkeit: Sicherungen werden so abgelegt, dass sie für einen festgelegten Zeitraum nicht verändert oder gelöscht werden können (WORM-Speicher, Objektsperren), oder physisch vom Netz getrennt gehalten.
- Getrennte Zugriffswege: Die Backup-Infrastruktur nutzt eigene Konten und Rechte, getrennt von den produktiven Administrationszugängen, damit ein übernommenes Produktivkonto die Sicherungen nicht erreicht.
- Integritätsprüfung: regelmäßige, möglichst automatisierte Prüfung, dass die Sicherungen vollständig und unverändert sind.
- Vertiefung: Vorgehen und Bausteine in NIST SP 1800-11 (Wiederherstellung nach Ransomware) und SP 1800-25 (Schutz der Datenbestände); Notfallplanung in NIST SP 800-34 Rev. 1; Anforderungen an Sicherung und Tests in C5:2026 OPS-06 bis OPS-09.
- Wirksamkeitstest: Aus einer unveränderlichen Kopie wird ein vollständiger Wiederherstellungstest durchgeführt; zusätzlich wird mit einem produktiven Administrationskonto versucht, eine Sicherung zu verändern oder zu löschen — dies muss verwehrt werden.

5. Umsetzungsbeispiele

Beispiel A — Entwicklungs-/Plattform-Kontext. Ein Anbieter sichert seine Systeme bisher regelmäßig, doch alle Sicherungen sind über dieselben Administrationszugänge erreichbar wie die produktiven Systeme. Ein Angreifer, der einen solchen Zugang übernimmt, könnte auch die Sicherungen verschlüsseln. Der Anbieter legt deshalb mindestens eine Kopie unveränderlich ab, sodass sie für einen festen Zeitraum nicht gelöscht oder verändert werden kann, und trennt die Verwaltung der Backup-Infrastruktur von den produktiven Zugängen. Vierteljährlich spielt er die Daten testweise vollständig zurück und dokumentiert das Ergebnis. So bleibt selbst bei einer weitreichenden Kompromittierung eine saubere Datenbasis erhalten.

Beispiel B — Allgemeine Abteilung. Eine Organisation ohne eigene IT-Entwicklung sichert ihre Dateien und Postfächer über einen Cloud-Dienst. Sie stellt sicher, dass mindestens eine Kopie unveränderlich aufbewahrt wird (viele Dienste bieten dafür eine Aufbewahrungs- oder Sperrfunktion) und dass die Verwaltung dieser Sicherung nicht an denselben Konten hängt wie die tägliche Administration. Einmal im Quartal stellt sie stichprobenhaft Dateien aus der Sicherung

wieder her und hält fest, dass es funktioniert hat. Damit erreicht sie den Kern des Schutzes auch ohne eigene Backup-Infrastruktur.

6. Reifegrad-Pfad (kumulativ)

- Initial: Backups vorhanden; Wiederherstellungstests jährlich.
- Defined: 3-2-1-1-0-Prinzip umgesetzt; Wiederherstellungstests vierteljährlich.
- Managed: unveränderliche Backups mit getrennten Zugriffswegen; automatisierte Integritätsprüfung.

7. Messung und Audit-Nachweis

- Kennzahl: Erfolgsquote der vierteljährlichen Wiederherstellungstests aus unveränderlichen Backups (Zielwert: 100 %).
- Nachweis: Protokolle der Wiederherstellungstests, Nachweis der Unveränderlichkeit (Aufbewahrungs- und Sperreinstellungen), getrennte Rechtevergabe der Backup-Infrastruktur, Berichte der Integritätsprüfung.
- Prüflogik nach ISO/IEC 27005: dokumentiert? — Richtlinie und Testprotokolle; verantwortlich? — IT-/Backup-Verantwortliche; Häufigkeit? — vierteljährliche Tests, fortlaufende Integritätsprüfung; Toleranz? — kein geschäftskritischer Datenbestand ohne eine unveränderliche oder getrennte Kopie und ohne bestandenen Wiederherstellungstest.

8. Typische Fehler

- Es gibt Backups, aber sie sind über dieselben Zugänge erreichbar wie die produktiven Systeme — ein Angreifer verschlüsselt beides.
- Sicherungen werden zwar erstellt, aber die Wiederherstellung wird nie vollständig getestet; im Ernstfall fehlen Teile oder der Vorgang dauert zu lange.
- „Unveränderlich" ist nur konfiguriert, aber nicht überprüft; eine zu kurze Sperrfrist macht die Kopie angreifbar.
- Die Sicherung umfasst nicht alle benötigten Daten (etwa Konfigurationen oder Schlüssel), sodass eine vollständige Wiederherstellung scheitert.

9. Abgrenzung, Restrisiko und Verweise

- Abgrenzung: betrifft Sicherung und Wiederherstellung; die durchgehende Verfügbarkeit über Redundanz behandelt A.8.14, die Wiederanlaufplanung A.5.29/A.5.30.
- Restrisiko: eine vom Netz getrennte Kopie ist naturgemäß weniger aktuell; eine sehr langsame oder selten getestete Wiederherstellung kann im Ernstfall dennoch zu Ausfallzeiten führen.
- Wirksamkeitsgrenze: Unveränderlichkeit schützt die Kopie, nicht die Wiederherstellung. Der Retention-Lock muss außerhalb der Kontroll-Ebene des Angreifers liegen — eigene Berechtigungen, getrennte Identität, kein Lösch- oder Verkürzungsrecht für kompromittierte Admin-Konten; entscheidend ist die regelmäßig getestete Wiederherstellung gegen ein definiertes RTO, nicht die bloße Existenz unveränderlicher Kopien.
- ISO/IEC 27002:2022: A.8.13 Sicherung von Informationen; A.8.14 Redundanz von informationsverarbeitenden Einrichtungen; A.5.29 Informationssicherheit bei Störungen; A.5.30 IKT-Bereitschaft für Business-Continuity; A.5.33 Schutz von Aufzeichnungen.
- Framework: C5:2026 OPS-06 bis OPS-09; DORA Art. 12; NIS2-DVO Nr. 4.1; NIST SP 800-34 Rev. 1; NIST SP 1800-11 und 1800-25 (Data Integrity / Ransomware-Recovery, NCCoE); Industriepraxis 3-2-1-1-0.

MHC-09 — AI-gestütztes Security-Testing in der Pipeline

Auf einen Blick

- Operativer Nutzen: Findet Schwachstellen schon vor der Auslieferung — automatisch bei jeder Änderung — und entzieht dem Angreifer das Zeitfenster zwischen Bekanntwerden einer Lücke und ihrem Schließen.
- Betroffene Richtlinie: Richtlinie zur sicheren Softwareentwicklung.
- Abhängigkeiten: setzt eine automatisierte Build-/CI-Pipeline voraus; nutzt die Stücklisten aus MHC-02 für den Komponenten-Abgleich.
- Aufwandstreiber: hängt von Zahl und Aufbau der Pipelines und der Menge des erzeugten Codes ab. Der konkrete Aufwand richtet sich nach den Ressourcen der Organisation.
- Primäre Kennzahl: Zeit vom Bekanntwerden einer aktiv ausgenutzten Schwachstelle (KEV) bis zum ausgerollten Patch; ergänzend Anteil der Pipelines mit Sicherheitsprüfung und Pre-Merge-Block.
- Berührte Standards (ohne Erfüllungsanspruch): ISO/IEC 27002 A.8.28/A.8.29; NIST SP 800-218 (SSDF) und SP 800-218A (KI); OWASP ASVS; C5:2026 OPS-25; CRA Annex I.

1. Control-Statement

Sicherheitsprüfungen laufen automatisch in der Entwicklungs-Pipeline: Code, Abhängigkeiten und laufende Anwendung werden bei jeder Änderung geprüft; schwerwiegende Funde blockieren die Übernahme. KI-generierter Code wird als eigene Risikokategorie zusätzlich geprüft.

2. Zweck und Bedrohungsbezug

Früher blieb zwischen dem Bekanntwerden einer Schwachstelle und einem funktionierenden Angriff Zeit zum Patchen. KI verkürzt dieses Fenster drastisch — teils auf Stunden. Wer Schwachstellen erst spät findet, ist zu langsam. Hinzu kommt: KI-Programmierhilfen erzeugen regelmäßig unsichere Muster (fest eingebaute Zugangsdaten, fehlende Prüfungen von Eingaben, unsichere Voreinstellungen, veraltete Bausteine). Schutzziel ist, Schwachstellen so früh wie möglich — schon bei jeder Änderung — automatisch zu finden und zu beheben, bevor sie in den Betrieb gelangen.

3. Organisatorische Umsetzung (CISO-Perspektive)

- Richtlinie: Die Entwicklungsrichtlinie schreibt automatische Sicherheitsprüfungen in der Pipeline vor, das Blockieren schwerwiegender Funde sowie die gesonderte Prüfung KI-generierten Codes.
- Verantwortlichkeiten: Im Statement of Applicability zugewiesen (Vorlage MRIS Anhang H); KI-generierter Code wird als eigene Risikokategorie geführt.
- Risikoanalyse und SoA: Schwellen für blockierende Funde und der Umgang mit aktiv ausgenutzten Schwachstellen (KEV) werden festgelegt. Behandeltes Risiko: ausnutzbare Schwachstellen, die unentdeckt in den Betrieb gelangen (Anbindung an die Risikobewertungs-Brücke, MRIS Anhang F).
- Prozesse: geregelter Ablauf für blockierende Funde und für KEV-Fälle (beschleunigtes Beheben); regelmäßige Auswertung der Prüfergebnisse; Penetrationstests mit auf die aktuelle Bedrohungslage zugeschnittenen Szenarien.

4. Technische Umsetzung (für IT-Fachleute)

- Prüfklassen in der Pipeline: statische Codeanalyse (SAST), dynamische Prüfung der laufenden Anwendung (DAST) und Prüfung der Abhängigkeiten (SCA) gegen Schwachstellendatenbanken — automatisch bei jeder Änderung.

- Blockieren (Pre-Merge-Block): schwerwiegende Funde (hoch/kritisch) ab einer festgelegten Verlässlichkeit verhindern die Übernahme; aktiv ausgenutzte Schwachstellen (KEV) blockieren sofort.
- Laufender Betrieb: regelmäßige (möglichst tägliche) Schwachstellen-Scans auch der bereits laufenden Systeme, nicht nur des neuen Codes.
- KI-Code als eigene Kategorie: zusätzliche Prüfregeln gegen die typischen Schwächen KI-generierten Codes (z. B. über Pre-Commit-Prüfungen); regelmäßige Auswertung mit Trendverfolgung; KI-generierter Code wird im Statement of Applicability als eigene Risikokategorie ausgewiesen.
- Vertiefung: sichere Entwicklungspraktiken in NIST SP 800-218 (SSDF), für KI-Modellentwicklung ergänzt durch SP 800-218A; Prüfkriterien für Anwendungssicherheit in OWASP ASVS; Scan-Anforderungen in C5:2026 OPS-25.
- Wirksamkeitstest: In einem Testzweig wird bewusst eine bekannte unsichere Stelle eingebaut (z. B. eine fest eingetragene Zugangskennung); die Pipeline muss den Fund melden und die Übernahme blockieren.
- Wirksamkeitsgrenze: AI-gestütztes Testing ersetzt nicht die fachliche Bewertung. Es übersieht insbesondere Logik-, Berechtigungs- und Architekturschwächen — also gerade Befunde mit hohem Impact — und erzeugt False Positives wie False Negatives. „Tool lief“ ist nicht „Schwachstelle blockiert“: kritische Findings, blockierende Pipeline-Entscheidungen und akzeptierte Ausnahmen brauchen fachliche Triage, dokumentierte Entscheidung und nachvollziehbare Freigabe.

5. Umsetzungsbeispiele

Beispiel A — Entwicklungs-/Plattform-Kontext. Ein Software-Anbieter prüft Sicherheit bisher erst spät, kurz vor einer Freigabe. Bei einer neu bekannten Schwachstelle ist er deshalb regelmäßig zu langsam. Er baut Sicherheitsprüfungen direkt in seine Pipeline ein: Code, Abhängigkeiten und laufende Anwendung werden bei jeder Änderung automatisch geprüft, und schwerwiegende Funde verhindern die Übernahme. Für aktiv ausgenutzte Schwachstellen gibt es einen beschleunigten Weg, sie sofort zu schließen. Da seine Entwickler KI-Programmierhilfen nutzen, ergänzt er Prüfregeln gegen die typischen Fehler solcher Werkzeuge und wertet diese Funde gesondert aus. So werden Lücken gefunden, bevor sie ausgeliefert werden.

Beispiel B — Allgemeine Abteilung. In einer Fachabteilung bauen Mitarbeitende mit Low-Code-Werkzeugen und KI-Programmierhilfen kleine eigene Anwendungen und Automatisierungen. Ihnen ist oft nicht bewusst, dass KI-generierter Code unsichere Muster enthalten kann. Die Organisation legt deshalb fest, dass solche selbst gebauten Lösungen nicht ungeprüft produktiv gehen: Alles, was über einfache Hilfsmittel hinausgeht, durchläuft eine fachliche und sicherheitstechnische Prüfung, und für sensible Daten oder Schnittstellen ist die zentrale IT einzubinden. So bleibt der Nutzen der schnellen Eigenentwicklung erhalten, ohne dass ungeprüfter KI-Code unkontrolliert in den Betrieb gelangt.

6. Reifegrad-Pfad (kumulativ)

- Initial: SAST/DAST/SCA in der Pipeline.
- Defined: zusätzliche Prüfung KI-generierten Codes; Blockieren bei schwerwiegenden Funden.
- Managed: tägliche Schwachstellen-Scans der laufenden Systeme; gesonderte Auswertung des KI-Codes; beschleunigter Weg für aktiv ausgenutzte Schwachstellen (KEV).

7. Messung und Audit-Nachweis

- Kennzahl: Zeit vom Bekanntwerden einer aktiv ausgenutzten Schwachstelle (KEV) bis zum ausgerollten Patch (Richtwert: möglichst unter 24 Stunden); ergänzend Anteil der Pipelines mit Sicherheitsprüfung und Pre-Merge-Block.
- Nachweis: Pipeline-Konfiguration, Prüf- und Blockier-Protokolle, Auswertung der KI-Code-Funde, Berichte der Penetrationstests.
- Prüflogik nach ISO/IEC 27005: dokumentiert? — Entwicklungsrichtlinie, Pipeline-Konfiguration; verantwortlich? — Entwicklungs-/Produktverantwortliche; Häufigkeit? — bei jeder Änderung, Scans täglich; Toleranz? — keine Auslieferung mit offenen schwerwiegenden Funden.

8. Typische Fehler

- Die Prüfungen laufen zwar, blockieren aber nicht; schwerwiegende Funde werden gemeldet und trotzdem ausgeliefert.
- Es wird nur der neue Code geprüft, nicht die bereits laufenden Systeme; dort bekannt gewordene Schwachstellen bleiben offen.
- KI-generierter Code wird wie handgeschriebener behandelt; seine typischen Schwächen werden nicht gezielt gesucht.
- Die Schwellen sind so streng gesetzt, dass die Pipeline ständig blockiert; in der Folge werden Prüfungen umgangen statt behoben.

9. Abgrenzung, Restrisiko und Verweise

- Abgrenzung: betrifft das automatisierte Prüfen während der Entwicklung; die Transparenz der Komponenten behandelt MHC-02, das Erkennen aktiver Angriffe MHC-05.
- Restrisiko: automatisierte Prüfungen finden bekannte Muster, nicht jede neuartige oder logische Schwachstelle; Penetrationstests und manuelle Prüfungen bleiben ergänzend nötig.
- ISO/IEC 27002:2022: A.8.8 Handhabung von technischen Schwachstellen; A.8.25 Lebenszyklus einer sicheren Entwicklung; A.8.26 Anforderungen an die Anwendungssicherheit; A.8.28 Sichere Codierung; A.8.29 Sicherheitsprüfung bei Entwicklung und Abnahme; A.8.32 Änderungssteuerung.
- Framework: C5:2026 OPS-25 (mit Verschärfung OPS-25.01AS, tägliche Scans); NIST SP 800-218 (SSDF v1.1; Revision 1.2 im Entwurf) und SP 800-218A (sichere Entwicklung für generative KI); OWASP ASVS; CRA Annex I Teil II.

MHC-10 — Continuous Control Monitoring und Policy-as-Code

Auf einen Blick

- Operativer Nutzen: Verkürzt die Zeit bis zum Erkennen einer Sicherheits-Abweichung von Monaten (Audit-Takt) auf Minuten, weil Vorgaben laufend automatisch geprüft werden.
- Betroffene Richtlinie: Richtlinie zur Überwachung der Wirksamkeit und Einhaltung von Sicherheitsvorgaben.
- Abhängigkeiten: keine Vorbedingung; setzt maschinenlesbar formulierbare Vorgaben und Zugriff auf System-/Konfigurationsdaten voraus.
- Aufwandstreiber: hängt von Zahl und Art der zu prüfenden Vorgaben und der Heterogenität der Systemlandschaft ab. Der konkrete Aufwand richtet sich nach den Ressourcen der Organisation.
- Primäre Kennzahl: Anteil der im SoA geführten Controls, deren Einhaltung automatisiert (mindestens täglich) geprüft wird (Continuous-Monitoring-Coverage).

- Berührte Standards (ohne Erfüllungsanspruch): ISO/IEC 27002 A.5.35/A.5.36; C5:2026 COM-03/04; NIS2-DVO Nr. 2.2; DORA Art. 6(5); NIST SP 800-137.

1. Control-Statement

Die Einhaltung von Sicherheitsvorgaben wird laufend und automatisiert gegen festgelegte Soll-Zustände geprüft, statt nur in periodischen Audits. Vorgaben werden, wo möglich, als ausführbare Regeln formuliert (Policy-as-Code) und in die Bereitstellungs- und Betriebsprozesse eingebunden; Abweichungen lösen automatisch eine Reaktion aus.

2. Zweck und Bedrohungsbezug

Periodische Audits prüfen nur zu festen Zeitpunkten (etwa quartalsweise); in der Zeit dazwischen bleiben Abweichungen unentdeckt. KI-gestützte Angreifer nutzen genau diese Lücke: Sie schaffen — oder finden — eine Fehlkonfiguration und nutzen sie, lange bevor das nächste Audit stattfindet. Zudem zerlegen agentische Angreifer ihr Vorgehen in Schritte, die jeder für sich regelkonform aussehen. Eine laufende, automatisierte Prüfung erkennt Abweichungen sofort und bringt sie unmittelbar zur Reaktion. Schutzziel ist, die Zeitspanne zwischen einer Abweichung und ihrer Erkennung auf ein Minimum zu verkürzen.

3. Organisatorische Umsetzung (CISO-Perspektive)

- Richtlinie: Die Überwachungsrichtlinie legt fest, dass die Einhaltung wesentlicher Vorgaben laufend automatisiert geprüft wird und Abweichungen unmittelbar bearbeitet werden — ergänzend zu, nicht statt der ohnehin nötigen Audits.
- Verantwortlichkeiten: Pflege der Regeln und Auswertung werden im Statement of Applicability zugewiesen (Vorlage MRIS Anhang H); fachlich verantwortlich ist in der Regel die ISMS-Leitung.
- Risikoanalyse und SoA: Es wird festgelegt, welche Controls vorrangig automatisiert überwacht werden — abgeleitet aus Risikobewertung und Schutzbedarf. Behandeltes Risiko: lange unentdeckte Abweichungen vom Soll-Zustand (Anbindung an die Risikobewertungs-Brücke, MRIS Anhang F).
- Prozesse: definierter Ablauf, wie eine erkannte Abweichung automatisch zu einem Vorgang (Ticket) und zur Behebung führt; regelmäßige Pflege und Erweiterung des Regelbestands; Berichtswesen über Abdeckung und offene Abweichungen.

4. Technische Umsetzung (für IT-Fachleute)

- Policy-as-Code: Sicherheitsvorgaben werden als ausführbare Regeln formuliert (z. B. mit Open Policy Agent/Rego oder HashiCorp Sentinel) und versioniert wie Quellcode verwaltet.
- Einbindung an zwei Stellen: vor der Bereitstellung in der CI-Pipeline (Prüfung, bevor etwas in Betrieb geht) und im laufenden Betrieb (fortlaufende Prüfung des Ist-Zustands gegen die Soll-Vorgaben), möglichst mit Durchsetzung (Runtime-Enforcement).
- Soll-Zustände (Baselines): die geprüften Soll-Vorgaben werden definiert und gepflegt; maschinenlesbare Control-Kataloge (z. B. im OSCAL-Format der NIST) erleichtern automatisierte Prüfung und Nachweis.
- Reaktion: Abweichungen werden auf Echtzeit-Übersichten mit klaren Kennzahlen dargestellt und lösen automatisch einen Vorgang (Incident-Ticket) aus.
- Vertiefung: Aufbau eines fortlaufenden Überwachungsprogramms in NIST SP 800-137 (ISCM) und dessen Bewertung in SP 800-137A; maschinenlesbare Controls über NIST OSCAL.

- Wirksamkeitstest: In einem Testbereich wird bewusst eine Abweichung vom Soll-Zustand erzeugt (z. B. eine nicht erlaubte Konfiguration); die automatisierte Prüfung muss die Abweichung binnen Minuten melden und einen Vorgang auslösen.

5. Umsetzungsbeispiele

Beispiel A — Entwicklungs-/Plattform-Kontext. Ein Unternehmen prüft die Einhaltung seiner Sicherheitsvorgaben bisher im Quartalsrhythmus per Audit. Zwischen den Prüfungen fällt eine fehlerhafte Freigabe in einer Cloud-Umgebung erst Wochen später auf. Das Unternehmen formuliert seine wichtigsten Vorgaben deshalb als ausführbare Regeln und bindet sie sowohl in die Bereitstellung als auch in den laufenden Betrieb ein. Verstößt eine Konfiguration gegen eine Regel, erscheint dies sofort auf einer Übersicht und erzeugt automatisch einen Vorgang zur Behebung. Aus der vierteljährlichen Momentaufnahme wird so eine laufende Kontrolle.

Beispiel B — Allgemeine Abteilung. Eine Organisation ohne eigene Entwicklung nutzt vor allem Cloud-Dienste. Sie kann keine eigenen Regeln programmieren, aktiviert aber die in ihren Plattformen vorhandenen Funktionen zur fortlaufenden Konfigurationsprüfung (etwa einen Sicherheits- oder Compliance-Status, der dauerhaft mitläuft). Abweichungen von den empfohlenen Einstellungen werden dort laufend angezeigt; die Organisation legt fest, wer sie aufnimmt und behebt. So erhält sie eine kontinuierliche Prüfung über die Bordmittel ihrer Dienste, ohne selbst zu entwickeln.

6. Reifegrad-Pfad (kumulativ)

- Initial: Compliance-Übersichten mit manueller Pflege.
- Defined: Policy-as-Code in der CI-Pipeline; Echtzeit-Prüfungen gegen Soll-Zustände.
- Managed: Durchsetzung im Betrieb (Runtime-Enforcement); automatisch erzeugte Vorgänge bei Abweichungen.

7. Messung und Audit-Nachweis

- Kennzahl: Continuous-Monitoring-Coverage — Anteil der im SoA geführten Controls mit aktiver automatisierter Prüfung (mindestens täglich, mit Alarmierung bei Abweichung). Richtwerte: ≥ 60 % nach 12 Monaten, ≥ 80 % nach 24 Monaten.
- Nachweis: Regel-Repository (Policy-as-Code), Übersicht der automatisiert geprüften Controls, Protokolle ausgelöster Vorgänge bei Abweichungen.
- Prüflogik nach ISO/IEC 27005: dokumentiert? — Regelbestand und Abdeckungsübersicht; verantwortlich? — ISMS-Leitung; Häufigkeit? — fortlaufend (mindestens täglich); Toleranz? — wesentliche Controls ohne automatisierte Prüfung nur befristet und begründet.

8. Typische Fehler

- Die automatisierte Prüfung wird als Ersatz für Audits verstanden; tatsächlich ergänzt sie diese — die geforderten Überprüfungen bleiben nötig.
- Es werden Übersichten gebaut, aber Abweichungen lösen keine Reaktion aus; der Befund bleibt folgenlos.
- Nur leicht prüfbare Vorgaben werden automatisiert, die wesentlichen aber nicht; die Abdeckung wirkt hoch, deckt aber nicht das Wichtige ab.
- Die Soll-Zustände werden einmal definiert und nicht gepflegt, sodass die Prüfung an veralteten Vorgaben misst.

9. Abgrenzung, Restrisiko und Verweise

- Abgrenzung: betrifft die laufende Prüfung der Einhaltung von Vorgaben; das Erkennen aktiver Angriffe behandelt MHC-05, die automatisierte Reaktion MHC-11.

- Restrisiko: geprüft wird nur, was als Regel formuliert ist — unklare oder nicht maschinell prüfbare Vorgaben bleiben außen vor; richtlinienkonforme, aber bösartige Aktivität wird erst im Zusammenspiel mit verhaltensbasierter Erkennung (MHC-05) sichtbar.
- ISO/IEC 27002:2022: A.5.18 Zugangsrechte; A.5.35 Unabhängige Überprüfung der Informationssicherheit; A.5.36 Einhaltung von Richtlinien, Vorschriften und Normen für die Informationssicherheit; A.8.9 Konfigurationsmanagement.
- Framework: C5:2026 COM-03/COM-04; NIS2-DVO Nr. 2.2; DORA Art. 6(5); OWASP SAMM; NIST SP 800-137 und 800-137A (ISCM); NIST OSCAL (maschinenlesbare Controls); Policy-Engines Open Policy Agent (OPA/Rego), HashiCorp Sentinel.

MHC-11 — SOAR-basierte Tier-1-Automation und parallele Reaktions-Playbooks

Auf einen Blick

- Operativer Nutzen: Bringt die Reaktion auf eindeutige Vorfälle von Stunden auf Minuten, indem die schnellen, klaren Eindämmungsschritte automatisch ablaufen — Menschen konzentrieren sich auf die mehrdeutigen und schwerwiegenden Fälle.
- Betroffene Richtlinie: Richtlinie zur Behandlung von Sicherheitsvorfällen (Incident Response).
- Abhängigkeiten: setzt verlässliche Erkennung voraus (MHC-05 liefert die Signale); die Automatisierungsschicht selbst muss gehärtet werden.
- Aufwandstreiber: hängt davon ab, ob eine SOAR-Plattform selbst betrieben oder als Dienst bezogen wird, und von der Zahl der Playbooks. Der konkrete Aufwand richtet sich nach den Ressourcen der Organisation.
- Primäre Kennzahl: Zeit bis zur Eindämmung eines eindeutigen Vorfalls (Mean Time to Containment, MTTC).
- Berührte Standards (ohne Erfüllungsanspruch): ISO/IEC 27002 A.5.24/A.5.26; C5:2026 SIM-02/03; NIS2-DVO Nr. 3.5; DORA Art. 17; NIST SP 800-61 Rev. 3.

1. Control-Statement

Auf eindeutige Vorfälle reagiert die Organisation automatisch mit vordefinierten Abläufen (Playbooks), die die Eindämmung in Minuten einleiten; mehrdeutige oder schwerwiegende Fälle gehen gezielt an Menschen. Die Automatisierungsschicht selbst wird abgesichert. Für mehrere gleichzeitige Vorfälle ist die Organisation vorbereitet.

2. Zweck und Bedrohungsbezug

KI-gestützte Angriffe laufen in Maschinengeschwindigkeit ab: Vom ersten Zugriff bis zum Schaden vergehen oft nur Minuten. Eine rein menschliche Reaktionskette — Alarm sichten, bewerten, eskalieren, handeln — ist dafür zu langsam. Zugleich führen Angreifer zunehmend mehrere Angriffe parallel, um die Reaktion zu überlasten. Schutzziel ist, die schnellen, eindeutigen Eindämmungsschritte zu automatisieren, damit sie in Minuten greifen, und auf mehrere gleichzeitige Vorfälle vorbereitet zu sein. Wichtig: Die Automatisierung wird selbst zum Angriffsziel und muss entsprechend geschützt werden.

3. Organisatorische Umsetzung (CISO-Perspektive)

- Richtlinie: Die Incident-Response-Richtlinie legt fest, welche Vorfälle automatisch eingedämmt werden, wann an Menschen eskaliert wird und wie die Automatisierungsschicht geschützt ist.

- Verantwortlichkeiten: Im Statement of Applicability zugewiesen (Vorlage MRIS Anhang H); der Sicherheitsbetrieb wird auf Analyse und Vorfallsleitung ausgerichtet, nicht auf reine Alarmsichtung.
- Risikoanalyse und SoA: Welche Aktionen automatisch laufen dürfen, wird nach Eindeutigkeit und möglicher Schadwirkung festgelegt. Behandeltes Risiko: zu langsame Reaktion gegenüber maschinengeschwinden und parallelen Angriffen (Anbindung an die Risikobewertungs-Brücke, MRIS Anhang F).
- Prozesse: regelmäßige Übungen für drei bis fünf gleichzeitige Vorfälle (mindestens halbjährlich); Pflege der Playbooks; klare Eskalationswege; bei Bezug als Dienst eine vertraglich zugesicherte Reaktionszeit.

4. Technische Umsetzung (für IT-Fachleute)

- Automatisierte Eindämmung: eine SOAR-Plattform führt bei eindeutigen Signalen vordefinierte Playbooks aus (z. B. Sperren eines Kontos, Isolieren eines Systems); mehrdeutige Fälle werden zur menschlichen Prüfung weitergegeben.
- Vorbereitete Szenarien: Playbooks für typische Mythos-Lagen, etwa Massen-Datenabfluss, paralleler Mehrfach-Angriff, Lieferketten-Kompromittierung, Angriffswellen auf Anmeldungen.
- Härtung der Automatisierungsschicht (zentral): Die SOAR-Pipeline ist selbst eine Angriffsfläche. Schutzmaßnahmen: nur authentisierte, signierte Auslöser aus vertrauenswürdigen Quellen; vollständiger Nachweis jeder Playbook-Ausführung samt ausführender Identität; Begrenzung der Auslöserate je Playbook (gegen Fehlalarm-Fluten); menschliche Freigabe (Human-in-the-Loop) für Aktionen mit großer Schadwirkung (z. B. Massen-Kontosperre, Isolation größerer Netzbereiche, Zertifikatsrückruf); Schutz der Playbooks wie Quellcode (Versionierung, Code-Review, signierte Freigaben).
- Vertiefung: Aufbau und Lebenszyklus der Vorfallsbehandlung in NIST SP 800-61 Revision 3 (am CSF 2.0 ausgerichtet); Anforderungen in C5:2026 SIM-02/03 und NIS2-DVO Nr. 3.5.
- Wirksamkeitstest: Ein eindeutiges Angriffssignal wird in einer Testumgebung ausgelöst; das passende Playbook muss die Eindämmung automatisch und innerhalb der Zielzeit einleiten — und eine Aktion mit großer Schadwirkung darf nur nach menschlicher Freigabe erfolgen.

5. Umsetzungsbeispiele

Beispiel A — Entwicklungs-/Plattform-Kontext. Ein Unternehmen mit eigenem Sicherheitsbetrieb stellt fest, dass die Zeit vom Alarm bis zur ersten Gegenmaßnahme zu lang ist, weil jeder Schritt manuell läuft. Es automatisiert deshalb die eindeutigen Eindämmungsschritte: Bei klaren Signalen wird ein betroffenes Konto sofort gesperrt oder ein System automatisch isoliert, während mehrdeutige Fälle an die Analysten gehen. Damit die Automatisierung nicht selbst zur Schwachstelle wird, akzeptiert die Plattform nur authentisierte Auslöser, protokolliert jede Aktion vollständig und verlangt für weitreichende Eingriffe eine menschliche Freigabe. Zweimal im Jahr übt das Team mehrere gleichzeitige Vorfälle. Die Zeit bis zur Eindämmung sinkt von Stunden auf Minuten.

Beispiel B — Allgemeine Abteilung. Eine Organisation ohne eigenen 24/7-Betrieb kann eine solche Automatisierung nicht selbst aufbauen und pflegen. Sie bezieht die Erkennung und schnelle Reaktion deshalb als Dienst (Managed Detection & Response) und achtet bei der Auswahl auf eine vertraglich zugesicherte Reaktionszeit sowie auf nachvollziehbare, abgestimmte Eindämmungsmaßnahmen. Intern legt sie fest, wer im Ernstfall ansprechbar ist und welche weitreichenden Aktionen vorab freigegeben sind. So erhält sie eine schnelle, geübte Reaktion, ohne eine eigene Plattform zu betreiben.

6. Reifegrad-Pfad (kumulativ)

- Initial: Reaktions-Playbooks dokumentiert; manuelle Ausführung.
- Defined: SOAR mit teilautomatisierten Playbooks; Zeit bis zur Eindämmung unter 60 Minuten.
- Managed: vollautomatische Eindämmung eindeutiger Vorfälle, Zeit unter 10 Minuten; halbjährliche Übungen mit mehreren gleichzeitigen Vorfällen — alternativ ein MDR-Dienst mit vertraglich zugesicherter Reaktionszeit.

7. Messung und Audit-Nachweis

- Kennzahl: Zeit bis zur Eindämmung eines eindeutigen Vorfalls (MTTC; Richtwert: unter 10 Minuten bei hoher Eindeutigkeit).
- Nachweis: Playbook-Definitionen und -Protokolle, Nachweis der Härtung der Automatisierungsschicht (Auslöser-Authentisierung, Audit-Trail, Freigaberegeln), Protokolle der Mehrfach-Vorfall-Übungen oder die Dienstleister-SLA.
- Prüflogik nach ISO/IEC 27005: dokumentiert? — Incident-Response-Richtlinie, Playbooks; verantwortlich? — Sicherheitsbetriebs-/SOC-Leitung; Häufigkeit? — fortlaufend, Übungen mindestens halbjährlich; Toleranz? — keine automatische Aktion mit großer Schadwirkung ohne menschliche Freigabe.

8. Typische Fehler

- Es wird breit automatisiert, ohne die Automatisierungsschicht zu schützen; nicht authentisierte Auslöser oder Fehlalarm-Fluten lösen ungewollte Eindämmungen aus.
- Weitreichende Aktionen laufen ohne menschliche Freigabe; ein Fehlalarm sperrt dann viele Konten oder isoliert ganze Netzbereiche.
- Es gibt Playbooks, aber keine Übung für mehrere gleichzeitige Vorfälle; im Ernstfall überlastet ein paralleler Angriff die Reaktion.
- Die Playbooks werden nicht wie Quellcode gepflegt (keine Versionierung, kein Review); Fehler und Manipulationen bleiben unbemerkt.

9. Abgrenzung, Restrisiko und Verweise

- Abgrenzung: betrifft die Reaktion auf erkannte Vorfälle; das Erkennen selbst behandelt MHC-05, die laufende Compliance-Prüfung MHC-10.
- Restrisiko: Die Automatisierung kann zum Ziel werden — Angreifer können Auslösebedingungen erkunden, gezielt Fehlalarme provozieren oder Eindämmungsaktionen gegen die eigene Organisation richten; ohne die beschriebene Härtung ist die Maßnahme nicht wirksam.
- Wirksamkeitsgrenze: SOAR ersetzt nicht belastbare Detection, Asset-Kontext und Incident-Governance. Automatisierung wirkt erst, wenn Auslöser, Datenqualität, Kritikalitätslogik, Freigaberegeln und Rückfallpfade definiert und getestet sind; Aktionen mit potenziell hoher Schadwirkung sehen standardmäßig menschliche Freigabe, Simulation oder gestufte Ausführung vor.
- ISO/IEC 27002:2022: A.5.5 Kontakt mit Behörden; A.5.24 Planung und Vorbereitung der Handhabung von Informationssicherheitsvorfällen; A.5.25 Beurteilung und Entscheidung über Informationssicherheitsereignisse; A.5.26 Reaktion auf Informationssicherheitsvorfälle.
- Framework: C5:2026 SIM-02/03; NIS2-DVO Nr. 3.5; DORA Art. 17; NIST SP 800-61 Revision 3 (Vorfallsbehandlung, am CSF 2.0 ausgerichtet).

MHC-12 — Threat-Led Penetration Testing mit Mythos-Szenarien

Auf einen Blick

- Operativer Nutzen: Prüft, ob die Schutzmaßnahmen gegen realistische, KI-typische Angriffsszenarien tatsächlich wirken — nicht nur, ob sie dokumentiert sind — und deckt so falsch-positive Sicherheitsannahmen auf.
- Betroffene Richtlinie: Richtlinie zu Sicherheitstests und unabhängiger Überprüfung.
- Abhängigkeiten: keine Vorbedingung; prüft die Wirksamkeit der übrigen Controls (u. a. MHC-05, MHC-11) unter realistischen Bedingungen.
- Aufwandstreiber: hängt stark vom gewählten Format ab (vollwertiges TLPT mit externem Red-Team vs. kostengünstigere Purple-Team- und Simulationsformate). Der konkrete Aufwand richtet sich nach den Ressourcen der Organisation.
- Primäre Kennzahl: Anteil der festgestellten Schwachstellen, die innerhalb des vereinbarten Zeitfensters behoben werden (Closure-Rate).
- Berührte Standards (ohne Erfüllungsanspruch): ISO/IEC 27002 A.5.35/A.8.29; DORA Art. 26/27 mit TLPT-RTS; TIBER-EU; C5:2026 OPS-22; NIS2-DVO Nr. 3.5.5.

1. Control-Statement

Die Wirksamkeit der Schutzmaßnahmen wird regelmäßig durch bedrohungsgeleitete Tests (Threat-Led Penetration Testing) geprüft, die realistische, auf die aktuelle Lage zugeschnittene Angriffsszenarien nachstellen — einschließlich KI-typischer Vorgehensweisen. Zwischen den Zyklen finden gemeinsame Übungen von Angriff und Abwehr (Purple Team) statt.

2. Zweck und Bedrohungsbezug

Dokumentierte Schutzmaßnahmen sind nicht dasselbe wie wirksame Schutzmaßnahmen. Unter KI-gestützten Angriffen — die heute nachweislich eigenständig Schwachstellen finden und ausnutzen — zeigt sich erst im realistischen Test, ob die Abwehr tatsächlich greift. Klassische Penetrationstests mit festem Standardumfang bilden die neuen Vorgehensweisen oft nicht ab. Schutzziel ist, mit realistischen, bedrohungsgeleiteten Szenarien zu prüfen, ob die Schutzmaßnahmen gegen reale Angriffsmuster wirken, und so falsch-positive Sicherheitsannahmen aufzudecken.

3. Organisatorische Umsetzung (CISO-Perspektive)

- Richtlinie: Die Testrichtlinie schreibt regelmäßige bedrohungsgeleitete Tests mit realistischen Szenarien vor sowie Purple-Team-Übungen zwischen den Zyklen.
- Verantwortlichkeiten: Im Statement of Applicability zugewiesen (Vorlage MRIS Anhang H); für die höchste Reife wird die fristgerechte Behebung der Funde gegenüber der Leitung berichtet.
- Risikoanalyse und SoA: Umfang und Szenarien werden aus der Bedrohungslage und den kritischen Funktionen abgeleitet; das geeignete Format wird nach Schutzbedarf und Mitteln gewählt. Behandeltes Risiko: Schutzmaßnahmen, die nur auf dem Papier wirken (Anbindung an die Risikobewertungs-Brücke, MRIS Anhang F).
- Prozesse: geregelte Beauftragung und Durchführung (bei vollwertigem TLPT mit getrennten Anbietern für Bedrohungsanalyse und Red-Team); Nachhalten und fristgerechtes Schließen der Funde; regelmäßige Wiederholung (mindestens jährlich für kritische Funktionen).

4. Technische Umsetzung (für IT-Fachleute)

- Bedrohungsgeleitete Tests: externe Red-Teamer stellen reale Angriffsketten nach, abgeleitet aus aktueller Bedrohungsanalyse — auf die Organisation zugeschnitten, gegen die produktiven Systeme.
- Mythos-Szenarien aufnehmen: z. B. KI-gestütztes Spear-Phishing mit gefälschter Stimme (Deepfake-Voice), in Mikroschritte zerlegte Angriffsketten, parallele Mehrfach-Angriffe, Lieferketten-Kompromittierungen, Missbrauch von Cloud-Zugängen über übernommene Dienstkonten.
- Purple Team zwischen den Zyklen: Angriffs- und Abwehrseite arbeiten gemeinsam und ordnen das Geübte den Techniken nach MITRE ATT&CK zu, um die Erkennungsabdeckung (MHC-05) gezielt zu verbessern.
- Abgestufte Formate: vollwertiges TLPT nach TIBER-EU-Methodik (aufwendig, für regulierte/kritische Stellen) oder kostengünstigere Formate — Purple-Team-Übungen mit ATT&CK-Szenarien, quelloffene Angriffssimulation (z. B. Caldera, Atomic Red Team) und Breach-and-Attack-Simulation-Plattformen (z. B. AttackIQ, SafeBreach, Picus).
- Vertiefung: TLPT-Vorgehen in der DORA-TLPT-RTS (Delegierte Verordnung (EU) 2025/1190) und im TIBER-EU-Rahmenwerk der EZB (seit Februar 2025 auf DORA ausgerichtet, mit verpflichtendem Purple Teaming); Penetrationstest-Anforderungen in C5:2026 OPS-22.
- Wirksamkeitstest: In einem bedrohungsgeleiteten Szenario (z. B. eine in Mikroschritte zerlegte Angriffskette) wird geprüft, ob die vorhandene Erkennung und Reaktion den Angriff tatsächlich bemerkt und stoppt — ein unbemerkter Durchlauf ist der Befund.

5. Umsetzungsbeispiele

Beispiel A — Entwicklungs-/Plattform-Kontext. Ein Unternehmen lässt jährlich einen Standard-Penetrationstest durchführen, der jedes Mal ähnlich abläuft. Neue, KI-typische Vorgehensweisen werden dabei nicht geprüft. Das Unternehmen geht deshalb zu bedrohungsgeleiteten Tests über: Externe Red-Teamer stellen reale Angriffsketten nach — etwa ein in viele kleine Schritte zerlegtes Vorgehen oder ein Phishing mit gefälschter Stimme — und prüfen gegen die echten Systeme, ob Erkennung und Reaktion greifen. Zwischen den Tests üben Angriffs- und Abwehrseite gemeinsam und schließen erkannte Lücken in der Erkennung. So wird sichtbar, welche Schutzmaßnahmen wirklich wirken und welche nur dokumentiert sind.

Beispiel B — Allgemeine Abteilung. Eine kleinere Organisation kann ein vollwertiges, aufwendiges Testprogramm nicht stemmen. Sie wählt deshalb ein kostengünstigeres, aber wirksames Format: regelmäßige gemeinsame Übungen von Angriff und Abwehr anhand realistischer Szenarien, unterstützt durch quelloffene Angriffssimulation oder eine Simulationsplattform, die typische Angriffstechniken automatisiert nachstellt. Die Ergebnisse zeigen konkret, wo die Erkennung nachgebessert werden muss. So prüft auch sie die Wirksamkeit ihrer Schutzmaßnahmen, angepasst an ihre Mittel.

6. Reifegrad-Pfad (kumulativ)

- Initial: jährliche Penetrationstests mit Standardumfang.
- Defined: bedrohungsgeleitete Tests mit Mythos-Szenarien; Purple-Team-Übungen zwischen den Zyklen.
- Managed: fortlaufende Angriffssimulation, Simulationsplattform produktiv; fristgerechte Behebung der Funde (Closure-Rate ≥ 90 % im vereinbarten Zeitfenster).

7. Messung und Audit-Nachweis

- Kennzahl: Closure-Rate — Anteil der festgestellten Schwachstellen, die innerhalb des vereinbarten Zeitfensters behoben werden (Richtwert: ≥ 90 %); ergänzend Häufigkeit und Realitätsnähe der Tests.
- Nachweis: Testberichte mit Szenarien und Funden, Behebungsnachweise, ATT&CK-Zuordnung der Purple-Team-Übungen.
- Prüflogik nach ISO/IEC 27005: dokumentiert? — Testrichtlinie, Testberichte; verantwortlich? — Sicherheitsleitung; Häufigkeit? — mindestens jährlich für kritische Funktionen, Purple Team dazwischen; Toleranz? — keine offenen kritischen Funde über das vereinbarte Zeitfenster hinaus.

8. Typische Fehler

- Es wird zwar getestet, aber mit immer gleichem Standardumfang; realistische, neue Angriffsmuster bleiben außen vor.
- Funde werden berichtet, aber nicht fristgerecht behoben; der Test bestätigt nur Bekanntes, ohne die Lage zu verbessern.
- Der Test prüft nur die Technik, nicht die Erkennung und Reaktion; ob ein Angriff bemerkt würde, bleibt offen.
- Das Format wird überdimensioniert (teures TLPT, wo ein Purple-Team-Format genügt) oder unterdimensioniert (oberflächlicher Scan, wo realistische Szenarien nötig wären).

9. Abgrenzung, Restrisiko und Verweise

- Abgrenzung: betrifft die Überprüfung der Wirksamkeit; das automatisierte Testen während der Entwicklung behandelt MHC-09, das laufende Erkennen MHC-05.
- Restrisiko: ein Test ist eine Momentaufnahme eines gewählten Szenarios; nicht geprüfte Wege bleiben offen, und die Lage ändert sich laufend — daher die regelmäßige Wiederholung und die Purple-Team-Übungen dazwischen.
- Formatabgrenzung: MHC-12 verlangt die Prüfung realistischer Angriffsketten, nicht zwingend ein DORA-TLPT im engen regulatorischen Sinn. Für DORA-pflichtige Einheiten bleibt das geregelte DORA-TLPT (RTS (EU) 2025/1190) verbindlich; alle anderen erreichen das Ziel auch über bedrohungsgeleitete Penetrationstests, Purple-Team-Übungen, Angriffssimulationen oder szenariobasierte Kontrolltests.
- ISO/IEC 27002:2022: A.5.35 Unabhängige Überprüfung der Informationssicherheit; A.8.29 Sicherheitsprüfung bei Entwicklung und Abnahme.
- Framework: DORA Art. 26/27 mit der TLPT-RTS (Delegierte Verordnung (EU) 2025/1190, anwendbar seit Juli 2025); TIBER-EU-Rahmenwerk der EZB (seit Februar 2025 auf DORA ausgerichtet); C5:2026 OPS-22; NIS2-DVO Nr. 3.5.5; Angriffssimulation u. a. mit Caldera, Atomic Red Team, AttackIQ, SafeBreach, Picus.

MHC-13 — AI-Agent-Governance und Harness-Sicherheit

Auf einen Blick

- Operativer Nutzen: Bringt selbst betriebene KI-Agenten unter Kontrolle — begrenzter Zugang, lückenlose Nachvollziehbarkeit, menschliche Freigabe vor riskanten Aktionen — und schließt damit eine sonst unkontrollierte Angriffsfläche.
- Betroffene Richtlinie: neue Richtlinie zur Nutzung von KI-Agenten; zusätzlich Zugriffssteuerungs- und Beschaffungsrichtlinie.
- Abhängigkeiten: setzt die Identitätsgrundlage aus MHC-04 voraus; berührt MHC-09 (unsicherer AI-Code) und das Shadow-AI-Inventar (A.5.9).

- Aufwandstreiber: hängt von Zahl und Reichweite der eingesetzten Agenten und Anbindungen ab; spätere Ausbaustufen setzen geeignete, im eigenen Umfeld verfügbare Werkzeuge voraus. Der konkrete Aufwand richtet sich nach den Ressourcen der Organisation.
- Primäre Kennzahl: Anteil der produktiven Agenten mit dokumentiertem Bedrohungsmodell und begrenztem Zugang sowie die Abdeckung der Protokollierung.
- Berührte Standards (ohne Erfüllungsanspruch): u. a. ISO/IEC 42001 Annex A, NIST AI RMF 1.0, OWASP-Bedrohungstaxonomien; je nach Einsatzbereich auch der EU AI Act.

1. Control-Statement

Produktiv eingesetzte AI-Agenten (Coding-Agenten, agentische Abläufe) werden wie privilegierte Systeme behandelt: begrenzter Schadensradius je Agent, eine auf die nötigen Rechte begrenzte Identität, eine lückenlose und nachvollziehbare Protokollierung mit benanntem menschlichen Verantwortlichen und Not-Aus, eine Freigabeliste für agentische Komponenten (MCP-Server, IDE-Erweiterungen, Skills) sowie Code-Review für den Harness (das Steuerungsgerüst des Agenten).

Mindestanforderungen je produktivem Agenten: technische, capability-scoped Identität statt persönlicher Benutzerkonten oder pauschaler API-Schlüssel; ausdrücklich kein unbeschränkter Tool-Zugriff; manipulationssichere, überprüfbare Ausführungsprotokollierung — das Protokoll ist sicherheitsrelevanter Nachweis und an die zentrale, manipulationssichere Protokollierung aus A.8.15 (MHC-05) angebunden; dokumentierte Zweckbindung und genehmigter Tool-Scope je Funktion; definierte Freigabeschwellen für Aktionen mit hoher Schadwirkung. Die Identitätsgrundlage liefert MHC-04 (Workload-Identität), auf der die capability-scoped Begrenzung aufsetzt.

2. Zweck und Bedrohungsbezug

AI-Agenten verbinden die Schlussfolgerung eines Sprachmodells mit Werkzeugausführung, dauerhaftem Speicher und mehrstufiger Planung. Dadurch entsteht eine privilegierte Angriffsfläche außerhalb der etablierten Sicherheitsmaßnahmen (Mythos-Ready, CSA/SANS/OWASP, April 2026: „Unmanaged AI Agent Attack Surface", Einstufung CRITICAL). Zwei Risikoseiten: erstens überprivilegierte oder unsichere Agenten im eigenen Betrieb; zweitens die Lieferkette — übernommene MCP-Server, IDE-Erweiterungen oder Skills. Belegt durch reale Vorfälle: EchoLeak (CVE-2025-32711, unbemerkter Datenabfluss über Prompt Injection in Microsoft 365 Copilot) und CurXecute (CVE-2025-54135/-54136, Codeausführung über die MCP-Anbindung der Coding-IDE Cursor).

3. Organisatorische Umsetzung (CISO-Perspektive)

- Richtlinie: Eine Richtlinie zur Nutzung von AI-Agenten wird erstellt. Sie regelt die erlaubten Einsatzzwecke, die Pflicht eines benannten menschlichen Verantwortlichen je produktivem Agenten, die ausschließliche Nutzung freigegebener Komponenten und die Pflicht zur menschlichen Freigabe vor nicht umkehrbaren Aktionen.
- Verantwortlichkeiten: Eine AI-Governance-Funktion wird benannt; der CISO ist rechenschaftspflichtig; der Vorstand wird einbezogen, weil die Freigabe produktiver Agenten Entscheidungen zur Risikoakzeptanz oberhalb des CISO-Mandats berührt (MRIS Anhang H, Kap. 10.5).
- Risikoanalyse und SoA: Vor der Produktivsetzung wird je Einsatzzweck eine Bedrohungs- und Risikobetrachtung erstellt und dokumentiert; daraus wird der Soll-Reifegrad abgeleitet (Anbindung an die Risikobewertungs-Brücke, MRIS Anhang F).
- Prozesse: (a) ein Freigabeprozess vor Produktivsetzung (dokumentierter Einsatzzweck, Risikobetrachtung, Rückfallplan); (b) ein Inventar der AI-Agenten und ihrer Komponenten;

(c) ein Verfahren zur lückenlosen Nachvollziehbarkeit mit benanntem Verantwortlichen und Aufbewahrung von mindestens 12 Monaten; (d) ein Verfahren zum schnellen Entzug übernommener Agenten.

- Beschaffung und Lieferanten: Anforderungen an externe agentische Komponenten (Herkunft, Update-Weg, Sicherheitsnachweise) werden festgelegt; Aufnahme nur über die Freigabeliste.
- Schulung: Entwicklungs- und Fachteams werden für die Risiken sensibilisiert, insbesondere für indirekte Prompt Injection.
- Harness als Code: Es wird festgelegt, dass die Steuerungsbestandteile eines Agenten (Anweisungen, Werkzeugdefinitionen, Regeldateien) denselben Freigabe- und Versionsregeln unterliegen wie Quellcode; die technische Umsetzung steht in Kapitel 4.

4. Technische Umsetzung (für IT-Fachleute)

- Begrenzte Identität: Jeder Agent läuft unter einer eigenen Workload-Identität (Anschluss an MHC-04) mit fein granularen, zeitlich begrenzten Rechten je Werkzeugaufruf (Lese-, Schreib-, Netzwerkrechte mit explizitem Scope); keine persönlichen Entwickler-Tokens.
- Begrenzung des Schadensradius: Rate-Limits je Agent-Identität; ein Circuit-Breaker, der bei ungewöhnlichen Aktionsfolgen abbricht; erzwungene Freigabeschwellen (Human-in-the-Loop) für nicht umkehrbare Aktionen (Produktiv-Deployment, Datenlöschung, Ausgaben oberhalb eines Schwellwerts).
- Protokollierung: Jeder Werkzeugaufruf sowie Datei- und Netzwerkzugriffe werden als wiederholbar nachvollziehbare Aufzeichnung protokolliert, gebunden an Sitzung, menschlichen Verantwortlichen und Datenquelle; revisionssichere Speicherung.
- Freigabeliste (Allowlist): technische Sperre nicht freigegebener MCP-Server und Erweiterungen; kontrollierte statt automatische Updates; Signaturprüfung, soweit verfügbar.
- Härtung gegen Prompt Injection: Trennung von Anweisungs- und Datenkanal, soweit technisch möglich; Ein- und Ausgangsfilterung; Behandlung der Werkzeug-Ausgaben als nicht vertrauenswürdige Eingabe (Schutz gegen missbrauchte Werkzeug-Berechtigung, „Confused Deputy"); Pen-Tests gegen Prompt Injection vor Produktivsetzung.
- Harness als Code: Anweisungen, Werkzeugdefinitionen, Retrieval-Pipelines und Eskalationslogik im Versionsverwaltungssystem, mit Code-Review-Pflicht, signierten Releases und automatisierten Prüfungen vor Produktivsetzung.
- Vertiefung: MCP-spezifische Kontrollen (Herkunftsprotokollierung, Sandboxing, Kontext-Isolation) im OWASP-Leitfaden „Practical Guide for Securely Using Third-Party MCP Servers"; Bedrohungstaxonomie in OWASP „Agentic AI — Threats and Mitigations" und in der OWASP Top 10 for Agentic Applications (ASI01–ASI10); AI-Management-Controls in ISO/IEC 42001:2023, Annex A; Adversarial-Robustheit der Agent-Modelle in NIST AI 100-2e2025.
- Wirksamkeitstest: Eine präparierte E-Mail oder ein präpariertes Ticket mit eingebettetem Befehl (Injection-Probe) wird eingespielt; der Agent darf die darin geforderte unzulässige Aktion (etwa externer Versand) nicht ausführen, und der Versuch muss im Protokoll erscheinen.

5. Umsetzungsbeispiele

Beispiel A — Entwicklungs-/Plattform-Kontext. In einer Organisation mit eigener Softwareentwicklung nutzen die Entwickler einen KI-Assistenten direkt in ihrer Arbeitsumgebung. Dieser Assistent ist mit mehreren internen Systemen verbunden — dem Ticketsystem, der Quellcodeverwaltung und einer internen Wissensdatenbank — und arbeitet dabei unter den persönlichen Zugängen der Entwickler. Das Risiko: In ein Ticket lässt sich ein verdeckter Befehl

einschleusen, der den Assistenten dazu bringt, Daten nach außen zu geben oder Quellcode zu verändern, ohne dass der Entwickler es bemerkt. Die Organisation führt zunächst ein vollständiges Verzeichnis aller eingesetzten Assistenten und ihrer Anbindungen, lässt nur geprüfte Anbindungen zu, gibt jedem Assistenten einen eigenen, eng begrenzten Zugang statt der vollen Entwicklerrechte und hält lückenlos fest, was er tut und wer ihn gestartet hat. In einer weiteren Ausbaustufe werden die Steuerungsvorgaben des Assistenten wie Programmcode geprüft und versioniert, regelmäßig auf Manipulierbarkeit getestet, und ein missbrauchter Zugang lässt sich binnen weniger Minuten entziehen.

Beispiel B — Allgemeine Abteilung (No-Code/SaaS). Eine Fachabteilung ohne eigene Entwicklung — etwa der Vertrieb — nutzt Microsoft 365 Copilot und richtet sich darin mit Bordmitteln einen eigenen Assistenten ein (oder verwendet ein selbst angelegtes GPT in ChatGPT Business). Dieser Assistent darf Postfach, Dateiablage und Kundendatenbank lesen und Nachrichten versenden, und zwar mit den vollen Rechten der Mitarbeiterin, die ihn eingerichtet hat. Das Risiko: Eine eingehende E-Mail oder ein Dokument kann einen verdeckten Befehl enthalten, der den Assistenten dazu bringt, vertrauliche Inhalte nach außen zu schicken — Genau dieses Angriffsmuster wurde im öffentlich dokumentierten EchoLeak-Fall für Microsoft 365 Copilot demonstriert. Hier liegen die Stellschrauben nicht in der Programmierung, sondern in den Verwaltungseinstellungen der Plattform: Die Abteilung erfasst, welche Assistenten es gibt und worauf sie zugreifen dürfen, lässt nur freigegebene Anbindungen zu, beschränkt die Rechte der Assistenten auf das Nötige statt der vollen Nutzerrechte, benennt für jeden produktiven Assistenten eine verantwortliche Person und schaltet die Protokollierung der Plattform ein. Für nicht umkehrbare Schritte — etwa den Versand nach außen oder das Löschen von Daten — wird eine ausdrückliche menschliche Bestätigung verlangt.

6. Reifegrad-Pfad (kumulativ)

- Initial: AI-Coding-Agenten dokumentiert; manuelles Inventar der MCP-Server und Erweiterungen.
- Defined: begrenzte Identitäten und lückenlose Protokollierung produktiv; Bedrohungsmodell je Einsatzzweck; Freigabeliste für Erweiterungen und MCP-Server.
- Managed: vollständiger Code-Review-Prozess für den Harness; automatisierte Prüfungen vor Produktivsetzung; Zeit bis zum Entzug unter 5 Minuten; regelmäßige Pen-Tests gegen Prompt Injection.
- Realismus: die meisten Organisationen starten bei Initial; realistisch 12–18 Monate bis Defined, 18–24 Monate bis Managed. Zuerst priorisieren: Inventar, Protokollierung und begrenzte Identitäten — diese drei adressieren den Großteil des Risikos. Pen-Tests des Harness und Prüfungen der Robustheit der Modelle folgen in späteren Ausbaustufen, abhängig von den im eigenen Umfeld verfügbaren Werkzeugen.

7. Messung und Audit-Nachweis

- Kennzahlen: Anteil der produktionsnahen Coding-Agenten mit dokumentiertem Bedrohungsmodell und festgelegtem Schadensradius (Zielwert: 100 %); Abdeckung der Protokollierung für Aktionen mit Schreib- oder Netzwerkrechten (Zielwert: 100 %); Zeit bis zum Entzug übernommener Agent-Identitäten (Zielwert: unter 5 Minuten).
- Nachweis: Agent-Inventar, Konfiguration der Freigabeliste, Protokolle, Prüfberichte vor Produktivsetzung, Bedrohungsmodell-Dokumente.
- Prüflogik nach ISO/IEC 27005: dokumentiert? — über Inventar und Bedrohungsmodelle; verantwortlich? — je Agent ein benannter Mensch; Häufigkeit? — fortlaufende Protokollierung, Prüfung vor jeder Produktivsetzung; Toleranz? — keine produktiven Agenten mit Schreib- oder Netzwerkrechten ohne Protokollierung.

8. Typische Fehler

- Agenten laufen unter persönlichen Entwickler-Konten — keine begrenzte Identität, keine saubere Nachvollziehbarkeit.
- MCP-Server und Erweiterungen ohne Freigabeliste und ohne Update-Kontrolle — automatische Updates öffnen einen Weg über die Lieferkette.
- Der Harness (Anweisungen, Regeldateien) wird als Konfiguration statt als Code behandelt — keine Freigabe, keine Versionsverwaltung.
- Menschliche Freigabe nur dem Namen nach (Bestätigungsdialog ohne echte Prüfmöglichkeit) — die zu weite Handlungsfreiheit bleibt bestehen.
- Das Ziel „Managed" wird pauschal angesetzt, obwohl die dafür nötigen Werkzeuge und Verfahren im eigenen Umfeld noch nicht etabliert sind.

9. Abgrenzung, Restrisiko und Verweise

- Abgrenzung: deckt Governance und Sicherheit der selbst betriebenen Agenten ab; die Grundlage der Identität liefert MHC-04; unsicherer, mit AI generierter Code wird in MHC-09 behandelt; das Inventar nicht freigegebener AI-Werkzeuge (Shadow AI) steht in A.5.9 (MRIS-Erweiterung).
- Restrisiko: Prompt Injection ist nicht vollständig lösbar, solange Anweisungs- und Datenkanal nicht trennbar sind; Erkennung ist nicht Verhinderung; geeignete Werkzeuge zur Härtung sind nicht in jedem Umfeld verfügbar.
- ISO/IEC 27002:2022: A.5.16 Identitätsmanagement; A.8.27 Sichere Systemarchitektur und Entwicklungsgrundsätze.
- AI-spezifisch: ISO/IEC 42001:2023, Annex A (AI-Management-Controls über das Statement of Applicability); NIST AI RMF 1.0; NIST AI 100-2e2025 (Adversarial Machine Learning).
- Bedrohungsmodelle: OWASP Top 10 for LLM Applications 2025 (LLM01 Prompt Injection; LLM06 Excessive Agency — mit den Ursachen zu weite Funktionalität, zu weite Rechte, zu weite Autonomie; LLM03 Supply Chain); OWASP Top 10 for Agentic Applications (ASI01 Agent Goal Hijack, ASI02 Tool Misuse, ASI03 Identity & Privilege Abuse); OWASP „Agentic AI — Threats and Mitigations"; OWASP „Practical Guide for Securely Using Third-Party MCP Servers"; MITRE ATLAS (AML.T0051 LLM Prompt Injection, AML.T0053 LLM Plugin Compromise, AML.T0086 Exfiltration via AI Agent Tool Invocation, AML.T0110 AI Agent Tool Poisoning).

Glossar

- **3-2-1-1-0:** Backup-Faustregel (drei Kopien, zwei Medientypen, eine offsite, eine unveränderlich/offline, null Fehler im Test).
- **AAL2 / AAL3:** Schutzniveaus für Authentisierung nach NIST SP 800-63B; AAL3 verlangt einen gerätegebundenen, nicht exportierbaren Schlüssel.
- **Admission-Controller:** Kontrollstelle der Plattform (z. B. in Kubernetes), die beim Deployment nur erlaubte, geprüfte Images zulässt.
- **Adversary Emulation / BAS:** Nachstellen realer Angriffstechniken, automatisiert über Breach-and-Attack-Simulation-Plattformen.
- **Air-gapped:** physisch vom Netz getrennte Sicherungskopie.
- **Baseline (Soll-Zustand):** definierter Soll-Zustand, gegen den der Ist-Zustand geprüft wird.
- **Blast Radius:** Schadwirkungsbereich — wie weit sich ein erfolgreicher Angriff ausbreiten kann.
- **Capability-Scoping:** Begrenzung eines Zugangs auf genau die Rechte und Fähigkeiten, die für die jeweilige Aufgabe nötig sind.
- **Conditional Access:** Zugriffsregeln, die bei jeder Anmeldung Nutzer, Gerät und Situation prüfen, bevor Zugriff gewährt wird.
- **Confidential Computing / TEE:** geschützte Ausführungsumgebung, die Daten auch während der Verarbeitung im Arbeitsspeicher verschlüsselt (z. B. Intel TDX, AMD SEV-SNP, ARM CCA).
- **Container / Image / Registry:** ein Container ist eine standardisiert verpackte Anwendung; das Image ist seine unveränderliche Vorlage; eine Registry ist der Ablageort für Images.
- **Continuous Control Monitoring:** laufende, automatisierte Prüfung der Einhaltung von Vorgaben statt periodischer Audits.
- **CVE/NVD, OSV, EUVD:** Schwachstellendatenbanken (NVD: US-Datenbank; OSV: Open-Source-Schwachstellen; EUVD: EU-Datenbank der ENISA).
- **CycloneDX / SPDX:** die beiden etablierten SBOM-Formate (CycloneDX als ECMA-424, SPDX als ISO/IEC 5962 standardisiert).
- **Deepfake-Voice:** künstlich erzeugte, täuschend echte Stimme, z. B. für Spear-Phishing.
- **Eindämmung (Containment):** sofortige Maßnahme, die einen laufenden Vorfall begrenzt (z. B. Konto sperren, System isolieren).
- **FIDO2 / WebAuthn:** offener Standard für phishing-resistente Anmeldung mit kryptografischer Bindung an die Adresse des Dienstes.
- **Harness:** Das Steuerungsgerüst eines KI-Agenten — Anweisungen, Werkzeugdefinitionen, Regeldateien und Eskalationslogik.
- **„heute abgreifen, später entschlüsseln":** Angriffsmuster, bei dem heute verschlüsselte Daten abgefangen und für eine spätere Entschlüsselung aufbewahrt werden.
- **Human-in-the-Loop:** verpflichtende menschliche Freigabe vor Aktionen mit großer Schadwirkung.
- **Hybride Kryptografie:** Kombination eines klassischen und eines quantensicheren Verfahrens während der Übergangszeit.
- **Image-Signatur (Sigstore/cosign):** kryptografische Signatur eines Images, die vor dem Start geprüft wird, um Manipulation auszuschließen.
- **KEV (Known Exploited Vulnerabilities):** Verzeichnis (CISA) der aktiv ausgenutzten Schwachstellen; löst beschleunigtes Patchen aus.
- **KI-generierter Code (Vibe-Coded):** mit KI-Programmierhilfen erzeugter Code, der als eigene Risikokategorie zu prüfen ist.
- **Kill-Chain-Korrelation:** Verknüpfung einzelner Ereignisse zu einer zusammenhängenden Angriffskette.

- **Living-off-the-Land:** Vorgehen, bei dem Angreifer bordeigene Systemwerkzeuge statt mitgebrachter Schadsoftware nutzen.
- **Mandant (Tenant) / Multi-Tenancy:** ein Kunde auf einer gemeinsam genutzten Plattform; Multi-Tenancy bezeichnet den Betrieb mehrerer Kunden auf gemeinsamer Infrastruktur.
- **MCP (Model Context Protocol):** Offene Schnittstelle, über die ein KI-Assistent mit externen Werkzeugen und Datenquellen verbunden wird.
- **MDR (Managed Detection & Response):** Bezug von Erkennung und schneller Reaktion als Dienst mit vertraglicher Reaktionszeit.
- **MITRE ATT&CK:** anerkannter Katalog realer Angriffstechniken, an dem sich die Erkennung ausrichtet.
- **ML-KEM / FIPS 203:** quantensicheres Verfahren für den Schlüsselaustausch (ersetzt z. B. RSA/ECDH).
- **mTLS (mutual TLS):** Verschlüsselte Verbindung, bei der sich beide Seiten gegenseitig per Zertifikat ausweisen — nicht nur eine Seite.
- **MTTC (Mean Time to Containment):** durchschnittliche Zeit bis zur Eindämmung eines Vorfalls.
- **OSCAL:** maschinenlesbares NIST-Format für Control-Kataloge und deren Nachweis.
- **Passkey:** auf FIDO2 beruhender Anmeldenachweis; gerätegebunden oder über mehrere Geräte synchronisierbar.
- **Playbook:** vordefinierter Reaktionsablauf für einen bestimmten Vorfallstyp.
- **Policy-as-Code:** Sicherheitsvorgaben als ausführbare, versionierte Regeln (z. B. OPA/Rego, Sentinel).
- **PQC (Post-Quantum-Kryptografie):** Verschlüsselungs- und Signaturverfahren, die auch künftigen Quantencomputern standhalten sollen.
- **Pre-Merge-Block:** automatische Sperre, die Code mit schwerwiegenden Funden nicht in den Hauptstand übernehmen lässt.
- **Prompt Injection (indirekt):** Ein in Inhalten (E-Mail, Dokument, Ticket) versteckter Befehl, der einen KI-Assistenten zu unerwünschtem Verhalten verleitet.
- **Red / Blue / Purple Team:** Angriffsseite, Abwehrseite und deren gemeinsame Übung.
- **Remote-Attestation:** Nachweis, dass eine geschützte Ausführungsumgebung echt und unverändert ist, bevor sie genutzt wird.
- **Runtime-Enforcement:** Durchsetzung von Vorgaben im laufenden Betrieb, nicht nur deren Prüfung.
- **SAST / DAST / SCA:** statische Codeanalyse, dynamische Prüfung der laufenden Anwendung, Prüfung der Abhängigkeiten.
- **SBOM (Software-Stückliste):** maschinenlesbares Verzeichnis aller Komponenten einer Software samt ihrer Abhängigkeiten.
- **Service Mesh:** Eine Infrastrukturschicht, die Verbindung und Absicherung zwischen Diensten automatisch übernimmt, ohne den Anwendungscode zu ändern.
- **SLSA / Build-Provenance:** Rahmenwerk und Nachweis, aus welchen Quellen und mit welchem Prozess ein Artefakt gebaut wurde.
- **SOAR:** Plattform, die definierte Reaktionsabläufe (Playbooks) automatisiert ausführt (Security Orchestration, Automation and Response).
- **SPIFFE / SPIRE:** Offener Standard (SPIFFE) und zugehörige Referenz-Implementierung (SPIRE), um jeder Software-Komponente eine eindeutige, technische Identität zu geben.
- **SSDF (Secure Software Development Framework):** NIST-Rahmen für sichere Softwareentwicklung (SP 800-218); für KI ergänzt durch SP 800-218A.
- **SVID:** Das von SPIFFE ausgestellte Identitätsdokument einer Komponente (als X.509-Zertifikat oder als Token), in der Regel mit kurzer Gültigkeit.

- **Threat-Hunting:** gezielte, hypothesengestützte Suche nach bislang unentdeckten Angriffen.
- **TIBER-EU:** Rahmenwerk der EZB für bedrohungsgeleitete Tests im Finanzsektor; seit Februar 2025 auf DORA ausgerichtet.
- **TLPT (Threat-Led Penetration Testing):** bedrohungsgeleiteter, realitätsnaher Test, der reale Angriffsszenarien nachstellt.
- **Transitive Abhängigkeit:** mittelbar enthaltener Baustein (eine Bibliothek, die ihrerseits weitere Bibliotheken nutzt).
- **UEBA:** verhaltensbasierte Erkennung anhand einer erlernten Normallinie für Nutzer und Systeme.
- **Unveränderlich (immutable) / WORM:** Speicherung, bei der Daten für einen festgelegten Zeitraum nicht verändert oder gelöscht werden können.
- **VEX / CSAF:** Formate zur Kommunikation der Ausnutzbarkeit von Schwachstellen, getrennt von der statischen SBOM.
- **ZTNA / Identity-Aware Proxy:** Zugang zu Anwendungen auf Basis der geprüften Identität von Nutzer und Gerät statt über die Zugehörigkeit zum Netz; Ersatz für klassisches VPN.

Mapping Implementation Guide ↔ MRIS 1.6, Kapitel 9

- **MHC-01 — Post-Quantum-Strategie und kryptografisches Inventar:** Katalog 9.2, Kompakt-Übersicht 9.4, Reifegrad-Stufen 9.5; Kennzahl: Kryptografisches-Inventar-Abdeckung.
- **MHC-02 — SBOM und Build-Provenance:** Katalog 9.2, Kompakt-Übersicht 9.4, Reifegrad-Stufen 9.5; Kennzahl: SBOM-Coverage.
- **MHC-03 — Phishing-resistente Multi-Faktor-Authentisierung:** Katalog 9.2, Kompakt-Übersicht 9.4, Reifegrad-Stufen 9.5; Lückencluster 6 (Identität/Zero Trust) in 9.3; Kennzahl: Phishing-resistent-MFA-Share.
- **MHC-04 — Workload-Identität und Zero-Trust-Netzwerkarchitektur:** Katalog 9.2, Kompakt-Übersicht 9.4, Reifegrad-Stufen 9.5; Lückencluster 6 in 9.3; Wirksamkeit über Implementierungs-Stage-Gates statt kontinuierlicher Kennzahl.
- **MHC-05 — Verhaltensbasierte Detection und Kill-Chain-Korrelation:** Katalog 9.2, Kompakt-Übersicht 9.4, Reifegrad-Stufen 9.5; Lückencluster 7 (Automatisierung/Resilienz) in 9.3; Kennzahl: ATT&CK-Coverage (strukturell), ergänzend Threat-Hunts.
- **MHC-06 — Container-Sicherheit und Confidential Computing:** Katalog 9.2, Kompakt-Übersicht 9.4, Reifegrad-Stufen 9.5; Lückencluster 3 (Container/Confidential Computing/Multi-Tenancy) in 9.3; Wirksamkeit über Implementierungs-Stage-Gates (Signatur-/Attestation-Abdeckung).
- **MHC-07 — Multi-Tenancy-Isolation mit nachweisbarer Trennung:** Katalog 9.2, Kompakt-Übersicht 9.4, Reifegrad-Stufen 9.5; Lückencluster 3 in 9.3; Wirksamkeit über Implementierungs-Stage-Gates (Ergebnis der Trennungstests).
- **MHC-08 — Unveränderliche Backups und Recovery-Validierung:** Katalog 9.2, Kompakt-Übersicht 9.4, Reifegrad-Stufen 9.5; Schnittmenge Lückencluster 3/4 in 9.3; Kennzahl: Restore-Test-Erfolgsquote.
- **MHC-09 — AI-gestütztes Security-Testing in der Pipeline:** Katalog 9.2, Kompakt-Übersicht 9.4, Reifegrad-Stufen 9.5; Lückencluster 7 in 9.3; Kennzahl: Patch-Latency für KEV-Listings.
- **MHC-10 — Continuous Control Monitoring und Policy-as-Code:** Katalog 9.2, Kompakt-Übersicht 9.4, Reifegrad-Stufen 9.5; Lückencluster 5 (Kontinuierliche Prüfung) in 9.3; Kennzahl: Continuous-Monitoring-Coverage.
- **MHC-11 — SOAR-basierte Tier-1-Automation und parallele Reaktions-Playbooks:** Katalog 9.2, Kompakt-Übersicht 9.4, Reifegrad-Stufen 9.5; Lückencluster 4/7 in 9.3; Kennzahl: Mean Time to Containment (MTTC).
- **MHC-12 — Threat-Led Penetration Testing mit Mythos-Szenarien:** Katalog 9.2, Kompakt-Übersicht 9.4, Reifegrad-Stufen 9.5; Lückencluster 5 in 9.3; Kennzahl: TLPT-Findings-Closure-Rate.
- **MHC-13 — AI-Agent-Governance und Harness-Sicherheit:** Katalog 9.2, Kompakt-Übersicht 9.4, Reifegrad-Stufen 9.5; Lückencluster 2 und 6 in 9.3; Wirksamkeit über Implementierungs-Stage-Gates (u. a. Mean Time to Revoke).

Zu allen MHC zusätzlich: RACI-Zuordnung in MRIS Anhang H, KPI-Definitionen in Anhang J/G, Einzelbewertung der berührten ISO-Controls in den MRIS-Kapiteln 4–6.