MYTHOS SYSTEMS

PERMANENT LIBRARY PUBLICATION

ENGINEERING STANDARDS LIBRARY / CYBERSECURITY AND NETWORK SECURITY

Penetration Testing and Ethical Hacking Standard

Defines authorization, scope, methodology, safety, evidence, reporting, remediation, and retest.

PERMANENT DOCUMENT ID
MYTHOS-SEC-0015

PUBLICATION
Candidate Standard

ASSURANCE SURFACE
6 Atomic Criteria / 6 Families

PERMANENT PUBLICATION IDENTITY

Penetration Testing and Ethical Hacking Standard

Defines authorization, scope, methodology, safety, evidence, reporting, remediation, and retest.

DOCUMENT ID
MYTHOS-SEC-0015

VERSION
1.0.0-candidate

STATUS
Candidate Standard

PUBLISHER
Mythos Systems

PUBLICATION DATE
2026-07-23

SCHEDULED REVIEW
2027-01-23

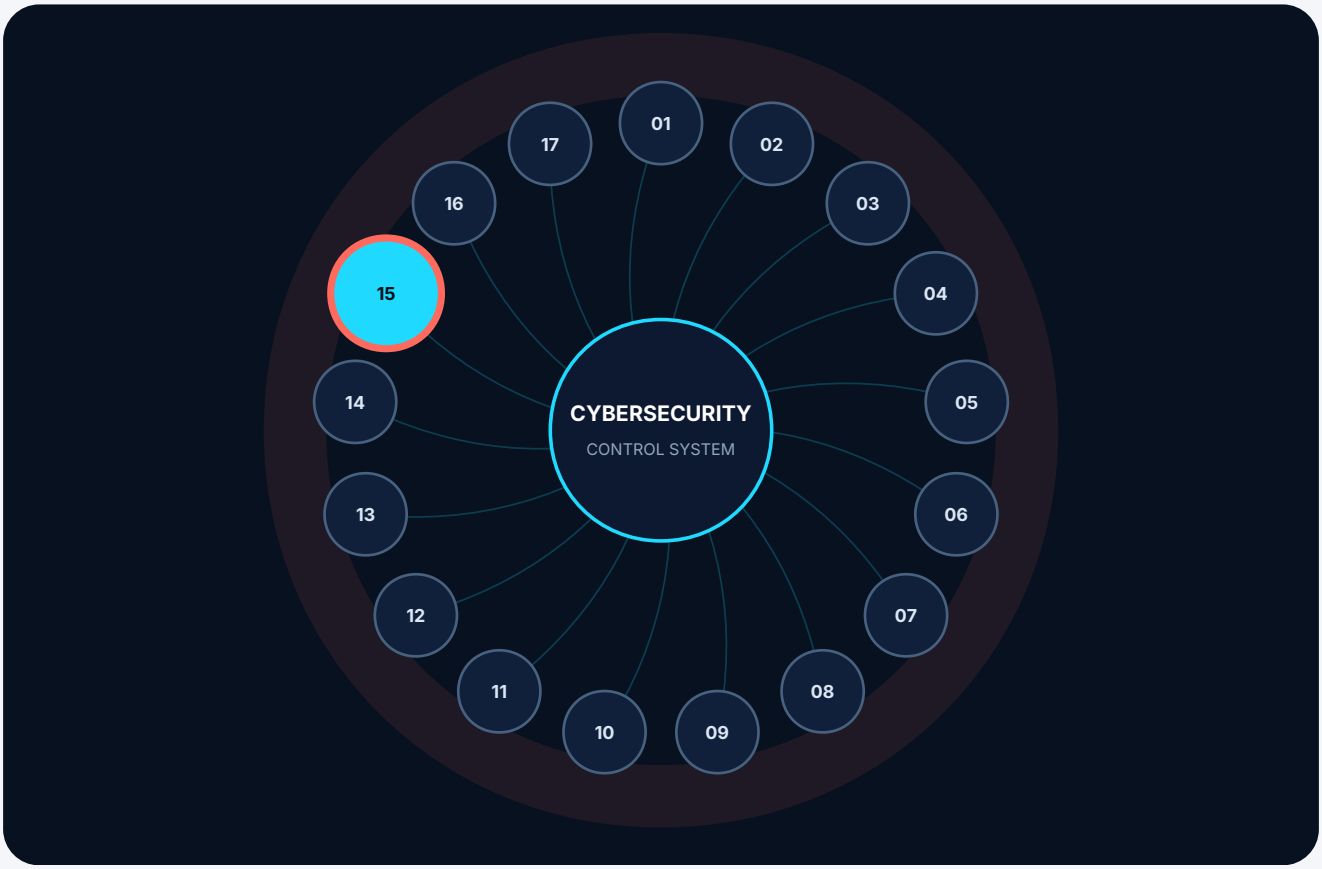
CLASSIFICATION
Public

6	6	4	1
ATOMIC CRITERIA	CAPABILITY FAMILIES	ASSESSMENT PROFILES	PERMANENT IDENTITY

Publication doctrine. This standard establishes durable requirements for its defined scope. Interpretation must preserve the permanent document identity, recorded evidence, and formal revision history.

DOMAIN CONTROL SYSTEM

MYTHOS-SEC-0015 inside the cybersecurity and network security standard constellation



Connected assurance

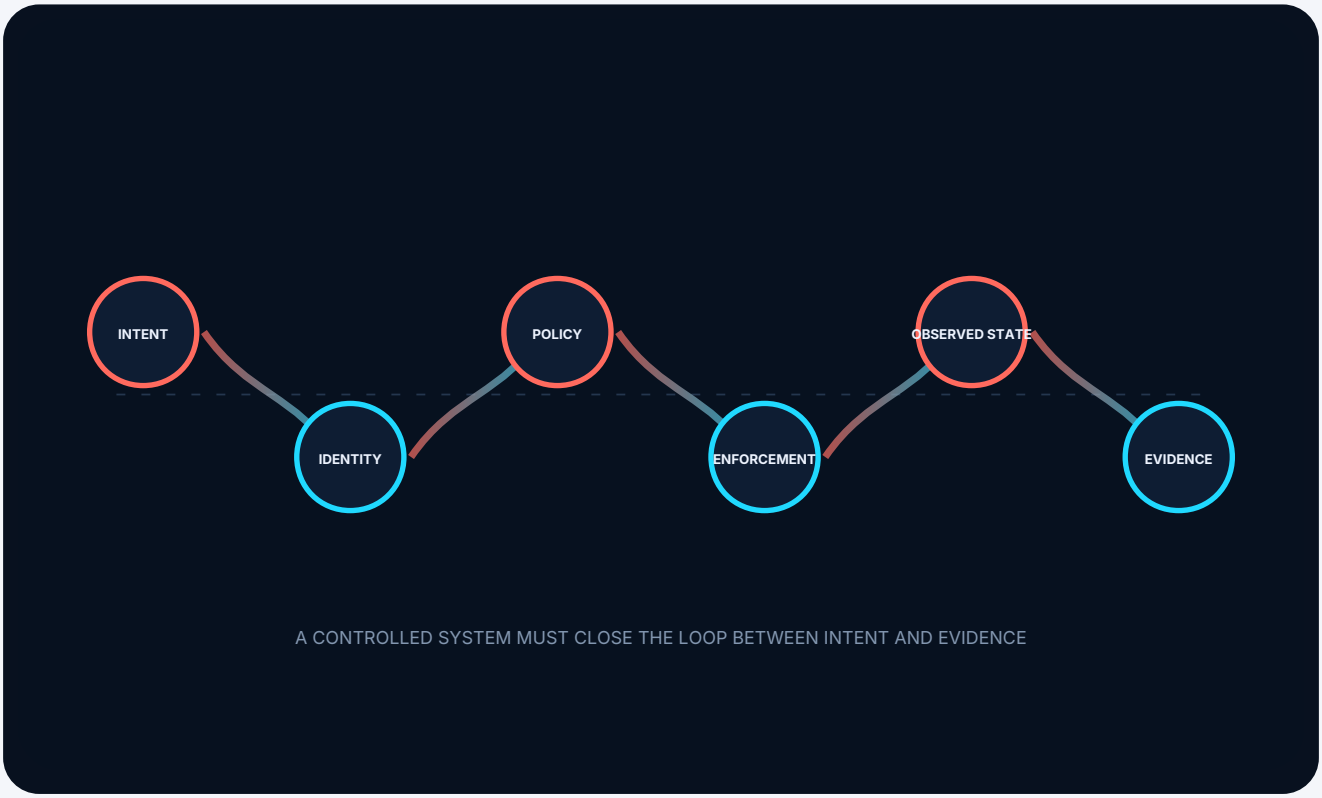
Architecture, policy, implementation, operations, and evidence are evaluated as one controlled system.

Specialization rule

Companion standards may add precision, but they cannot silently weaken this publication's requirements.

CLOSED-LOOP ARCHITECTURE

Intent becomes trustworthy only when observed state returns as evidence



State separation

Desired, approved, deployed, observed, historical, and simulated states must remain distinguishable.

Recoverable change

Material change requires validation, authorization, monitoring, containment, and a credible recovery path.

EVALUATION TOPOLOGY

The capability families and atomic criteria of MYTHOS-SEC-0015

Capability families

Families expose the major engineering surfaces and preserve the source weighting model.

Atomic criteria

Each criterion carries a question, maturity spectrum, evidence expectations, assessment method, and operational signals.

NAVIGATION

Contents

ENGINEERING DOCTRINE

0. Executive Mandate

The architecture of security validation has failed.

Organations treat penetration testing as an annual compliance checkbox. They hire external firms to run automated vulnerability scanners against their production endpoints, receive a 200-page PDF of low-severity findings, and consider the system "secure." When a real attacker breaches them a month later using a chained exploit that the scanner missed, they are mystified.

This standard exists because:

1. **Automated Scanning is Not Penetration Testing:** A scanner identifies known vulnerabilities. A penetration tester identifies *business logic flaws* and chains multiple low-severity vulnerabilities into a catastrophic breach. Relying on scanners provides a false sense of security.
2. **Compliance Does Not Equal Security:** Passing a compliance audit (e.g., SOC 2, FedRAMP) proves you have documentation; it does not prove your architecture can withstand a determined adversary. Penetration testing MUST be threat-model-driven, not compliance-driven.
3. **Unscoped Engagements are Reckless:** Ethical hacking without strict Rules of Engagement (RoE) and legal authorization is just a cyberattack. Systems MUST enforce hard boundaries, emergency stop mechanisms, and strict data handling protocols to prevent testers from causing accidental outages or exposing PII.
4. **Reports Must Be Reproducible:** A penetration test report that states "We achieved RCE" without providing the exact payload, execution context, and reproduction steps is useless to developers. Findings MUST include granular, step-by-step mechanics and remediation code.

This document establishes the **Mythos Penetration Testing Standard (MPTS)**, a comprehensive, evidence-based methodology for evaluating ethical hacking programs against formal methodology adherence, exploit realism, and operational safety.

ENGINEERING DOCTRINE

Part I. Scope and Applicability

1.1 In Scope

This standard covers the evaluation of:

- Penetration Testing Execution Standard (PTES) and OSSTMM methodologies
- Rules of Engagement (RoE) and legal authorization scoping
- White Box, Grey Box, and Black Box testing paradigms
- Adversary emulation and Red Teaming (MITRE ATT&CK)
- Exploit chaining and business logic validation
- Purple Teaming and reproducible remediation validation

1.2 Out of Scope

- General vulnerability management (covered in MYTHOS-SEC-0008)
- Defensive architecture rules (covered in MYTHOS-SEC-0001 through 0014)

1.3 Audience

- Penetration testers, red team operators, and ethical hackers
- CISOs and security directors managing testing programs
- DevSecOps teams consuming penetration test reports
- GRC teams evaluating compliance mandates
- Standards bodies seeking a reference ethical hacking methodology

---

ENGINEERING DOCTRINE

Part II. Definitions and Taxonomy

2.1 Ethical Hacking Dimensions

Dimension	Definition
Rules of Engagement (RoE)	The legally binding document defining the scope, timing, intensity, and boundaries of a penetration test.
Black Box Testing	A testing paradigm where the attacker has zero prior knowledge of the target architecture, simulating an external threat.
Grey Box Testing	A testing paradigm where the attacker is given partial knowledge (e.g., user-level credentials) to simulate an insider threat or lateral movement.
Exploit Chaining	The practice of linking multiple distinct vulnerabilities (e.g., Open Redirect + XSS + CSRF) to achieve a catastrophic outcome (e.g., Account Takeover).
Purple Teaming	A collaborative exercise where Red (offense) and Blue (defense) teams work in real-time to validate exploit mechanics and defensive detections simultaneously.

2.2 The Ethical Hacking Doctrine

> A scanner finds bugs; a human finds breaches. An exploit without scope is a crime. A report without reproduction steps is noise. If the test doesn't simulate real adversary behavior, it is just compliance theater.

ENGINEERING DOCTRINE

Part III. Evaluation Methodology

3.1 Scoring Model

Each capability is scored from 0 to 5.

Score	Meaning
0	Fails basic task; automated scans only, no RoE, compliance-driven
1	Manual testing exists, but ad-hoc, unscoped, and lacks reproduction
2	Functional testing, but relies on checklists rather than threat modeling
3	Solid production performance; PTES methodology, strict RoE, exploit chaining
4	Strong performance; Purple teaming, ATT&CK emulation, zero false positives
5	Best-in-class; sets the standard for mathematically verified adversary emulation

Weighting

Category	Weight	Rationale
Methodology & Scoping (PTES/OSSTMM)	20%	Unscoped testing is reckless and legally dangerous
Threat Realism & Exploit Chaining	20%	Single-vuln testing misses catastrophic breach paths
Rules of Engagement & Safety	15%	Testers must not cause production outages or data exposure
Reporting & Reproducibility	20%	Developers cannot fix what they cannot reproduce
Black/Grey/White Box Coverage	15%	Relying on one paradigm leaves blind spots
Remediation & Purple Team Validation	10%	Findings must be verified as fixed by the offense team

Total: 100%

A system MUST score at least 3.0 in Methodology/Scoping and Reporting to be considered a valid penetration test.

EVALUATION AND ASSURANCE SYSTEM

Capability claims are bounded by evidence, context, and reproducible assessment.

6 atomic criteria across 6 capability families define the evaluation surface of this standard.

4.1 / CAPABILITY FAMILY

Methodology and Scoping

SOURCE WEIGHT 20%

4.1.1
Framework Adherence

MYTHOS-SEC-0015-EVAL-4.1.1

Framework Adherence

FAMILY	SOURCE REFERENCE	WEIGHT CONTEXT	ASSESSMENT POSTURE
Methodology and Scoping	4.1.1	20%	Evidence-led

Is the penetration test governed by a formal, industry-standard methodology?

0
Ad-hoc testing based on tester intuition

1
Basic OWASP Top 10 checklist followed

2
PTES methodology followed, but scoping is too narrow (e.g., prod only, no CI/CD)

3
Full PTES/OSSTMM adherence; scope includes network, app, APIs, and CI/CD pipelines

4
Threat-model-driven scoping: tests are designed around specific adversary personas

5
Leading methodology: formally verified scope coverage, zero blind spots, mathematically complete attack surface mapping

ENGINEERING INTERPRETATION

This criterion evaluates whether framework adherence is governed as an operating capability rather than a one-time configuration. Assessment must distinguish intended design, approved implementation, observed behavior, historical evidence, and unresolved risk.

REQUIRED EVIDENCE

- approved architecture or policy record
- current configuration or platform export
- change and exception records
- monitoring, test, or assessment evidence

ASSESSMENT PROCEDURE

1. Examine authoritative design, policy, configuration, and lifecycle records.
2. Interview the accountable owner and operators responsible for the control.
3. Test a representative positive, negative, failure, and recovery scenario.

EXAMINE - INTERVIEW - TEST

COMMON FAILURE PATTERNS

- The control exists only in diagrams or policy text.
- Observed state differs from approved intent without a governed exception.
- Evidence cannot establish scope, time, owner, or operating effectiveness.

OPERATIONAL MEASURES

- coverage of in-scope assets or flows
- age and closure rate of unresolved findings
- percentage of changes passing pre-deployment validation
- time to detect and contain control failure

DECISION BOUNDARY

A maturity score is valid only for the assessed scope and evidence period. Documentation alone does not establish operating effectiveness.

4.2 / CAPABILITY FAMILY

Threat Realism and Exploit Chaining

SOURCE WEIGHT 20%

4.2.1
Adversary Emulation

MYTHOS-SEC-0015-EVAL-4.2.1

Adversary Emulation

FAMILY	SOURCE REFERENCE	WEIGHT CONTEXT	ASSESSMENT POSTURE
Threat Realism and Exploit Chaining	**4.2.1**	**20%**	**Evidence-led**

Does the test simulate a real, determined attacker rather than just running a scanner?

0
Automated scanner output handed to client

1
Manual validation of scanner findings, but no attempt to chain exploits

2
Basic chaining (e.g., SSRF to internal metadata endpoint)

3
Complex chaining: combining business logic flaws, low-severity vulns, and network pivots to achieve ATO/RCE

4
Full Red Team engagement: utilizing MITRE ATT&CK TTPs, social engineering, and physical access (if in scope)

5
Leading realism: formally verified attack paths, zero reliance on automated tools, mathematical proof of exploitability

ENGINEERING INTERPRETATION

This criterion evaluates whether adversary emulation is governed as an operating capability rather than a one-time configuration. Assessment must distinguish intended design, approved implementation, observed behavior, historical evidence, and unresolved risk.

REQUIRED EVIDENCE

- approved architecture or policy record
- current configuration or platform export
- change and exception records
- monitoring, test, or assessment evidence

ASSESSMENT PROCEDURE

1. Examine authoritative design, policy, configuration, and lifecycle records.
2. Interview the accountable owner and operators responsible for the control.
3. Test a representative positive, negative, failure, and recovery scenario.

EXAMINE - INTERVIEW - TEST

COMMON FAILURE PATTERNS

- The control exists only in diagrams or policy text.
- Observed state differs from approved intent without a governed exception.
- Evidence cannot establish scope, time, owner, or operating effectiveness.

OPERATIONAL MEASURES

- coverage of in-scope assets or flows
- age and closure rate of unresolved findings
- percentage of changes passing pre-deployment validation
- time to detect and contain control failure

DECISION BOUNDARY

A maturity score is valid only for the assessed scope and evidence period. Documentation alone does not establish operating effectiveness.

4.3 / CAPABILITY FAMILY

Rules of Engagement and Safety

SOURCE WEIGHT 15%

4.3.1
Operational Boundaries

MYTHOS-SEC-0015-EVAL-4.3.1

Operational Boundaries

FAMILY	SOURCE REFERENCE	WEIGHT CONTEXT	ASSESSMENT POSTURE
Rules of Engagement and Safety	**4.3.1**	**15%**	**Evidence-led**

Are strict legal and operational controls enforced during testing?

0
No RoE; tester has unrestricted access to production

1
Basic scope document, but no emergency stop mechanism

2
RoE defines timing and targets, but lacks data handling protocols

3
Strict RoE: defines scope, timing, emergency stop, and PII handling protocols

4
Real-time monitoring of tester actions; automated kill-switch if production impact detected

5
Leading safety: formally verified blast radius containment, zero ability to impact production availability, cryptographic auditing of all tester actions

ENGINEERING INTERPRETATION

This criterion evaluates whether operational boundaries is governed as an operating capability rather than a one-time configuration. Assessment must distinguish intended design, approved implementation, observed behavior, historical evidence, and unresolved risk.

REQUIRED EVIDENCE

- approved architecture or policy record
- current configuration or platform export
- change and exception records
- monitoring, test, or assessment evidence

ASSESSMENT PROCEDURE

1. Examine authoritative design, policy, configuration, and lifecycle records.
2. Interview the accountable owner and operators responsible for the control.
3. Test a representative positive, negative, failure, and recovery scenario.

EXAMINE - INTERVIEW - TEST

COMMON FAILURE PATTERNS

- The control exists only in diagrams or policy text.
- Observed state differs from approved intent without a governed exception.
- Evidence cannot establish scope, time, owner, or operating effectiveness.

OPERATIONAL MEASURES

- coverage of in-scope assets or flows
- age and closure rate of unresolved findings
- percentage of changes passing pre-deployment validation
- time to detect and contain control failure

DECISION BOUNDARY

A maturity score is valid only for the assessed scope and evidence period. Documentation alone does not establish operating effectiveness.

4.4 / CAPABILITY FAMILY

Reporting and Reproducibility

SOURCE WEIGHT 20%

4.4.1
Actionable Intelligence

MYTHOS-SEC-0015-EVAL-4.4.1

Actionable Intelligence

FAMILY	SOURCE REFERENCE	WEIGHT CONTEXT	ASSESSMENT POSTURE
Reporting and Reproducibility	4.4.1	20%	Evidence-led

Can the development team independently reproduce and fix the exploit?

0
Report states "Vulnerable to XSS" with no payload

1
Report includes the payload, but no reproduction steps

2
Report includes step-by-step reproduction, but lacks remediation code

3
Report includes exact payloads, HTTP requests, execution context, and specific remediation code (e.g., "Use DOMPurify here")

4
Report includes video/screen recording of the exploit chain

5
Leading reporting: formally verified reproduction steps, automated CI/CD test cases generated for the vulnerability, zero ambiguity in remediation

ENGINEERING INTERPRETATION

This criterion evaluates whether actionable intelligence is governed as an operating capability rather than a one-time configuration. Assessment must distinguish intended design, approved implementation, observed behavior, historical evidence, and unresolved risk.

REQUIRED EVIDENCE

- approved architecture or policy record
- current configuration or platform export
- change and exception records
- monitoring, test, or assessment evidence

ASSESSMENT PROCEDURE

1. Examine authoritative design, policy, configuration, and lifecycle records.
2. Interview the accountable owner and operators responsible for the control.
3. Test a representative positive, negative, failure, and recovery scenario.

EXAMINE - INTERVIEW - TEST

COMMON FAILURE PATTERNS

- The control exists only in diagrams or policy text.
- Observed state differs from approved intent without a governed exception.
- Evidence cannot establish scope, time, owner, or operating effectiveness.

OPERATIONAL MEASURES

- coverage of in-scope assets or flows
- age and closure rate of unresolved findings
- percentage of changes passing pre-deployment validation
- time to detect and contain control failure

DECISION BOUNDARY

A maturity score is valid only for the assessed scope and evidence period. Documentation alone does not establish operating effectiveness.

4.5 / CAPABILITY FAMILY

Black/Grey/White Box Coverage

SOURCE WEIGHT 15%

4.5.1
Perspective Diversity

MYTHOS-SEC-0015-EVAL-4.5.1

Perspective Diversity

FAMILY	SOURCE REFERENCE	WEIGHT CONTEXT	ASSESSMENT POSTURE
Black/Grey/White Box Coverage	4.5.1	15%	Evidence-led

Is the architecture tested from multiple threat perspectives?

0
Only Black Box testing (misses internal logic flaws)

1
Only White Box testing (code review, misses runtime config issues)

2
Grey Box testing only (simulates insider, misses external zero-days)

3
Combination of Black Box (external) and Grey Box (insider) testing

4
Full spectrum: Black, Grey, and White box testing mapped to specific threat actors

5
Leading coverage: formally verified attack surface mapping from all perspectives, zero blind spots

ENGINEERING INTERPRETATION

This criterion evaluates whether perspective diversity is governed as an operating capability rather than a one-time configuration. Assessment must distinguish intended design, approved implementation, observed behavior, historical evidence, and unresolved risk.

REQUIRED EVIDENCE

- approved architecture or policy record
- current configuration or platform export
- change and exception records
- monitoring, test, or assessment evidence

ASSESSMENT PROCEDURE

1. Examine authoritative design, policy, configuration, and lifecycle records.
2. Interview the accountable owner and operators responsible for the control.
3. Test a representative positive, negative, failure, and recovery scenario.

EXAMINE - INTERVIEW - TEST

COMMON FAILURE PATTERNS

- The control exists only in diagrams or policy text.
- Observed state differs from approved intent without a governed exception.
- Evidence cannot establish scope, time, owner, or operating effectiveness.

OPERATIONAL MEASURES

- coverage of in-scope assets or flows
- age and closure rate of unresolved findings
- percentage of changes passing pre-deployment validation
- time to detect and contain control failure

DECISION BOUNDARY

A maturity score is valid only for the assessed scope and evidence period. Documentation alone does not establish operating effectiveness.

4.6 / CAPABILITY FAMILY

Remediation and Purple Team Validation

SOURCE WEIGHT 10%

4.6.1
Fix Verification

MYTHOS-SEC-0015-EVAL-4.6.1

Fix Verification

FAMILY	SOURCE REFERENCE	WEIGHT CONTEXT	ASSESSMENT POSTURE
Remediation and Purple Team Validation	**4.6.1**	**10%**	**Evidence-led**

Is the fix validated by the offense team, not just the developers?

0
Client marks finding as "fixed" in a spreadsheet; no verification

1
Tester re-runs scanner to verify fix

2
Tester manually re-tests the specific vulnerability

3
Purple Team exercise: Red attacks, Blue defends, real-time validation of the fix

4
Automated regression tests added to CI/CD to prevent the vuln from reappearing

5
Leading validation: formally verified non-exploitability, mathematical proof that the attack vector is permanently closed

ENGINEERING INTERPRETATION

This criterion evaluates whether fix verification is governed as an operating capability rather than a one-time configuration. Assessment must distinguish intended design, approved implementation, observed behavior, historical evidence, and unresolved risk.

REQUIRED EVIDENCE

- approved architecture or policy record
- current configuration or platform export
- change and exception records
- monitoring, test, or assessment evidence

ASSESSMENT PROCEDURE

1. Examine authoritative design, policy, configuration, and lifecycle records.
2. Interview the accountable owner and operators responsible for the control.
3. Test a representative positive, negative, failure, and recovery scenario.

EXAMINE - INTERVIEW - TEST

COMMON FAILURE PATTERNS

- The control exists only in diagrams or policy text.
- Observed state differs from approved intent without a governed exception.
- Evidence cannot establish scope, time, owner, or operating effectiveness.

OPERATIONAL MEASURES

- coverage of in-scope assets or flows
- age and closure rate of unresolved findings
- percentage of changes passing pre-deployment validation
- time to detect and contain control failure

DECISION BOUNDARY

A maturity score is valid only for the assessed scope and evidence period. Documentation alone does not establish operating effectiveness.

LIFECYCLE AND ASSURANCE

Part V. The Mythos Penetration Testing Suite (MPTS)

Traditional security benchmarks measure scanner coverage. The MPTS evaluates methodology adherence, chaining realism, and report reproducibility.

5.1 Suite Composition

5.1.1 MPTS-Methodology

- **Scope Audit:** 100+ tests verifying the engagement scope includes APIs, CI/CD, cloud infrastructure, and network, not just the web UI.
- **RoE Compliance:** 50+ tests verifying the tester did not execute out-of-scope attacks or cause production impact.

5.1.2 MPTS-Reproduction

- **Payload Validation:** 100+ tests taking the report's reproduction steps and executing them in a clean environment to verify they work.
- **Chain Reconstruction:** 50+ tests verifying the reported exploit chain actually results in the claimed impact (e.g., ATO).

5.2 Execution Protocol

PUBLICATION BOUNDARY

Candidate standard. Permanent identity. Evidence before authority.

Approval requires designated technical, security, legal, accessibility, and governance review with recorded findings and dispositions.