MYTHOS SYSTEMS

PERMANENT LIBRARY PUBLICATION

ENGINEERING STANDARDS LIBRARY / CYBERSECURITY AND NETWORK SECURITY

Zero-Trust Policy, Authority Separation, and Capability Grant Standard

Defines continuous authorization, least privilege, separation of duties, capability grants, and evidence.

PERMANENT DOCUMENT ID
MYTHOS-SEC-0003

PUBLICATION
Candidate Standard

ASSURANCE SURFACE
5 Atomic Criteria / 5 Families

PERMANENT PUBLICATION IDENTITY

Zero-Trust Policy, Authority Separation, and Capability Grant Standard

Defines continuous authorization, least privilege, separation of duties, capability grants, and evidence.

policy-enforcement

DOCUMENT ID
MYTHOS-SEC-0003

VERSION
1.0.0-candidate

STATUS
Candidate Standard

PUBLISHER
Mythos Systems

PUBLICATION DATE
2026-07-23

SCHEDULED REVIEW
2027-01-23

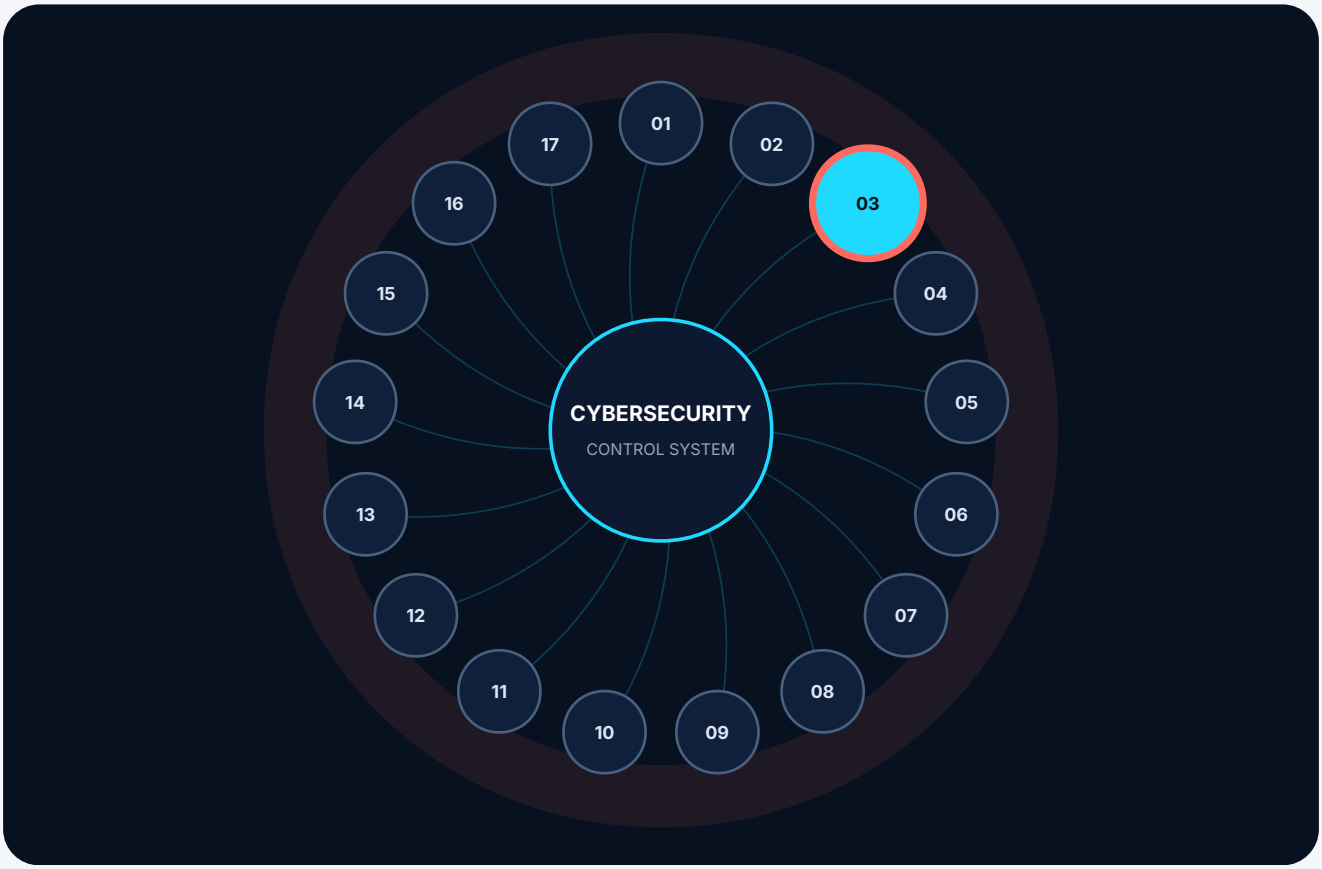
CLASSIFICATION
Public

5	5	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-0003 inside the cybersecurity and network security standard constellation



Connected assurance

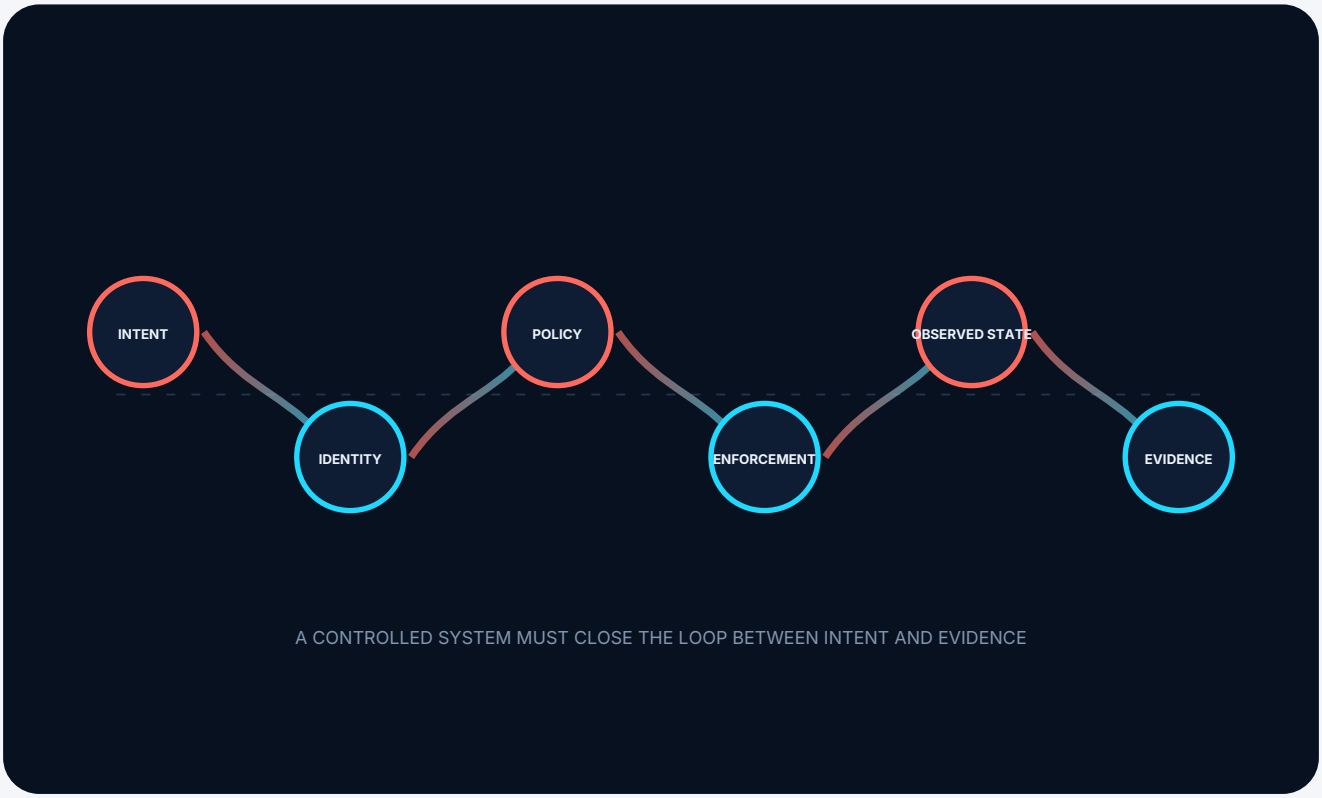
Architecture, policy, implementation, operations, and evidence are evaluated as one controlled system.

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-0003

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

Traditional evaluation of AI authorization is insufficient.

Current agentic frameworks treat the LLM as a trusted entity. The model is granted broad tool access, environment variables, and API keys, and it executes commands directly. This is a catastrophic security posture. An LLM is a probabilistic engine susceptible to prompt injection, sycophancy, and hallucination. Granting an LLM direct execution authority is equivalent to giving a junior engineer root access to production infrastructure based on a vibe-check.

This standard exists because:

1. **Models are Not Principals:** A model is not an identity. It cannot hold credentials. It cannot be trusted to enforce its own safety guidelines. Every consequential action MUST pass through a deterministic, policy-governed proxy.
2. **RBAC is Insufficient for Agents:** Role-Based Access Control (RBAC) assumes a static hierarchy. Agentic systems are dynamic: a parent agent spawns a child agent to execute a sub-task. The child MUST NOT inherit the parent's full permissions; it MUST receive an attenuated, scoped, and revocable capability grant.
3. **Policy Must Be Mathematically Verifiable:** "Do not delete production" written in a system prompt is a suggestion. Policy MUST be written in a formal language (Cedar) and evaluated by a deterministic engine outside the model's context.
4. **Secrets Must Never Enter the Context Window:** Passing an AWS API key into a prompt so the model can call an SDK is a data exfiltration vulnerability. Secrets MUST be brokered directly to the execution environment, never visible to the model.

This document establishes the **Mythos Zero-Trust & Authority Standard (MZTAS)**, a comprehensive, evidence-based methodology for evaluating the strict separation of AI reasoning from execution authority.

ENGINEERING DOCTRINE

Part I. Scope and Applicability

1.1 In Scope

This standard covers the evaluation of:

- Authority separation (Models propose, deterministic engines dispose)
- Policy-as-Code engines (Cedar, OPA) for agentic tool execution
- Capability grants and cryptographic attenuation (Macaroons)
- Zero-trust agent mesh architecture (preventing privilege escalation)
- Secret brokering without model visibility (JIT, scoped, zeroization)
- Hardware-backed roots of trust for capability tokens

1.2 Out of Scope

- General network security (covered in MYTHOS-NET-0004)
- Supply chain integrity (covered in MYTHOS-SEC-0002)

1.3 Audience

- Principal engineers and architects selecting agentic authorization infrastructure
- Security teams evaluating agentic attack surfaces and privilege escalation
- Platform engineering teams building internal agent platforms
- AI governance, risk, and compliance teams
- Standards bodies seeking a reference zero-trust AI methodology

---

ENGINEERING DOCTRINE

Part II. Definitions and Taxonomy

2.1 Authority Dimensions

Dimension	Definition
Authority Separation	The architectural principle that models propose actions, but a deterministic, independent policy engine authorizes them, and a deterministic executor performs them.
Capability Grant	A temporary, scoped, revocable permission issued to an agent to request a specific action.
Macaroon Attenuation	The cryptographic process of restricting a capability token's scope (e.g., reducing path access or setting an expiration) without contacting the issuer.
Zero-Trust Mesh	An architecture where no agent is trusted by default; all inter-agent communication is authenticated and authorized per-call.

2.2 The Zero-Trust Doctrine

> Models propose. Policy authorizes. Determinism executes. Evidence records. A model that holds a secret is a breach. An agent that inherits permissions is a liability.

ENGINEERING DOCTRINE

Part III. Evaluation Methodology

3.1 Scoring Model

Each capability is scored from 0 to 5.

Score	Meaning
0	Fails basic task; model has direct execution authority
1	Basic interception but policy is advisory, not enforced
2	Policy enforced but coupled to the framework, not pluggable
3	Solid production performance; pluggable policy, basic capabilities
4	Strong performance; full capability attenuation, zero-trust mesh
5	Best-in-class; sets the standard for agentic zero-trust

Weighting

Category	Weight	Rationale
Authority Separation	20%	Model-driven execution is the #1 cause of AI outages
Policy-as-Code Enforcement	20%	Unverifiable policies are suggestions
Capability Grants & Attenuation	15%	RBAC causes privilege escalation in multi-agent systems
Secret Isolation	15%	Secrets in prompts are exfiltration vectors
Zero-Trust Agent Mesh	10%	Implicit trust between agents is a lateral movement path
Hardware-Backed Roots	10%	Software tokens can be exfiltrated
Operational Lifecycle	10%	Policies must be governed like code

Total: 100%

A system MUST score at least 3.0 in Authority Separation to be considered for production use.

EVALUATION AND ASSURANCE SYSTEM

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

5 atomic criteria across 5 capability families define the evaluation surface of this standard.

4.1 / CAPABILITY FAMILY

Authority Separation

SOURCE WEIGHT 20%

4.1.1
Model Execution Isolation

MYTHOS-SEC-0003-EVAL-4.1.1

Model Execution Isolation

FAMILY	SOURCE REFERENCE	WEIGHT CONTEXT	ASSESSMENT POSTURE
Authority Separation	4.1.1	20%	Evidence-led

Does the framework enforce separation between model reasoning and execution authority?

0
Model output directly executes code or commands

1
Interception exists but execution is not policy-governed

2
Policy enforcement exists but is advisory (can be bypassed)

3
Policy enforcement is pluggable and enforced outside the model context

4
Full authority separation: models propose, independent engine authorizes, deterministic executor performs

5
Leading separation: formally verified policy engine, cryptographic handoff between proposal and execution

ENGINEERING INTERPRETATION

This criterion evaluates whether model execution isolation 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

Policy-as-Code Enforcement

SOURCE WEIGHT 20%

4.2.1
Policy Evaluation

MYTHOS-SEC-0003-EVAL-4.2.1

Policy Evaluation

FAMILY	SOURCE REFERENCE	WEIGHT CONTEXT	ASSESSMENT POSTURE
Policy-as-Code Enforcement	**4.2.1**	**20%**	**Evidence-led**

Is policy enforcement deterministic and formally verifiable?

0
No policy enforcement

1
Basic string matching or LLM self-policing

2
Standard RBAC checks

3
Pluggable policy engine (OPA/Cedar) with typed contracts

4
Statically analyzable authorization (Cedar) with formal logic

5
Leading policy: formally verified, mathematically analyzable, and cryptographically bound to the execution context

ENGINEERING INTERPRETATION

This criterion evaluates whether policy evaluation 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

Capability Grants and Attenuation

SOURCE WEIGHT 15%

MYTHOS-SEC-0003-EVAL-4.3.1

Delegation Control

FAMILY	SOURCE REFERENCE	WEIGHT CONTEXT	ASSESSMENT POSTURE
Capability Grants and Attenuation	**4.3.1**	**15%**	**Evidence-led**

Can a parent agent securely delegate permissions to a child agent without over-privileging?

0
Child agents inherit all parent permissions

1
Child agents get static roles

2
Basic scoping exists but cannot be revoked

3
Temporary, scoped capability grants with per-call validation

4
Cryptographic capability tokens (Macaroons) with attenuation

5
Leading delegation: cryptographic attenuation, per-delegation revocation, and zero-privilege-escalation guarantees

ENGINEERING INTERPRETATION

This criterion evaluates whether delegation control 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

Secret Isolation

SOURCE WEIGHT 15%

4.4.1
Secret Visibility

MYTHOS-SEC-0003-EVAL-4.4.1

Secret Visibility

FAMILY	SOURCE REFERENCE	WEIGHT CONTEXT	ASSESSMENT POSTURE
Secret Isolation	4.4.1	15%	Evidence-led

Can tools use secrets without exposing them to the model?

0
Secrets passed in prompts or environment variables visible to the model

1
Secrets loaded but not redacted from logs

2
Secrets brokered to execution environments without model visibility

3
Short-lived, scoped leases with automatic redaction

4
Hardware-backed secret brokering with zeroization

5
Leading isolation: zero-knowledge brokering, hardware-backed roots, and cryptographic audit of every secret access

ENGINEERING INTERPRETATION

This criterion evaluates whether secret visibility 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

Zero-Trust Agent Mesh

SOURCE WEIGHT 10%

4.5.1
Inter-Agent Trust

MYTHOS-SEC-0003-EVAL-4.5.1

Inter-Agent Trust

FAMILY	SOURCE REFERENCE	WEIGHT CONTEXT	ASSESSMENT POSTURE
Zero-Trust Agent Mesh	4.5.1	10%	Evidence-led

Are all agent-to-agent communications authenticated and authorized?

0
All agents share a global trust domain

1
Basic identity exists but no per-call authorization

2
Per-call authorization but no identity verification

3
Mutual TLS and per-call policy evaluation

4
Cryptographic identity and zero-trust mesh

5
Leading mesh: zero implicit trust, cryptographic identity, and formal verification of inter-agent isolation

ENGINEERING INTERPRETATION

This criterion evaluates whether inter-agent trust 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 Zero-Trust Suite (MZTS)

Traditional security benchmarks evaluate network firewalls. The MZTS evaluates agentic authorization boundaries.

5.1 Suite Composition

5.1.1 MZTS-Authority

- **Execution Isolation:** 100+ tests verifying the model cannot execute without deterministic policy approval.
- **Privilege Escalation:** 50+ tests attempting to bypass the policy engine via prompt injection.

5.1.2 MZTS-Capabilities

- **Attenuation:** 100+ tests verifying child agents cannot exceed their attenuated scope.
- **Revocation:** 50+ tests verifying revoked capabilities are instantly rejected.

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.