Product Requirements Document: Agent and MCP Control Plane

Field	Value
ID	prd-agent-control-plane-001
Version	1.1.0
Status	Draft
Created	2025-12-19
Updated	2025-12-19
Author(s)	John Wang
Tags	agent-governance, identity, security, spiffe, multi-tenancy, mcp

---

1. Executive Summary

1.1 Problem Statement

Organizations deploying AI agents face critical governance challenges: no standardized way to cryptographically identify AI agents, no mechanism to verify agent capabilities and security posture before granting access, agents often store API keys and OAuth tokens directly creating security risks, limited visibility into what agents access and on whose behalf, and no centralized management of expensive LLM API usage across agents.

1.2 Proposed Solution

The Agent and MCP Control Plane provides centralized governance infrastructure for AI agent and MCP server ecosystems, implementing cryptographic identity via SPIFFE/SPIRE, trust attestation via Know Your Agent (KYA) certification, just-in-time credential injection via a governance proxy, comprehensive audit logging of all agent and MCP server actions, and LLM usage management with model-level policies and cost tracking.

1.3 Expected Outcomes

- 100% of agents have cryptographically verifiable SPIFFE IDs
- Zero API keys stored in agent code or configuration
- 100% of API calls logged with full identity context
- Mean time to revoke access under 1 minute
- 100% LLM cost attribution to agent and user

1.4 Target Audience

Organizations deploying AI agents requiring enterprise-grade governance, identity, and security

1.5 Value Proposition

End-to-end agent identity, attestation, credential injection, and observability in a unified control plane - addressing a market gap where no existing solution provides this combination

---

2. Objectives and Goals

2.1 Business Objectives

ID	Objective	Rationale	Aligned With
bo-1	Capture leadership position in the $9-10B agent governance market by 2030	AI Agents market growing at 46.3% CAGR with significant governance gaps	Market opportunity analysis
bo-2	Achieve $250M ARR by Year 5 through enterprise customer acquisition	Conservative market capture of 1,000+ enterprise customers	SOM projections
bo-3	Differentiate from cloud IAM, zero trust, and API gateway competitors	No existing solution provides agent-native identity and governance	Competitive landscape analysis

2.2 Product Goals

ID	Goal	Rationale
pg-1	Implement zero-trust agent identity where every agent has a cryptographically verifiable identity	Foundation for all governance capabilities
pg-2	Achieve credential isolation where agents never possess API keys or OAuth tokens	Eliminates credential exposure risk

ID	Goal	Rationale
pg-3	Enable delegated access with scoped, time-limited user authorization to agents	Supports least-privilege access patterns
pg-4	Provide full auditability with every credential access logged with agent and user context	Required for compliance and incident response
pg-5	Implement cost governance with centralized control over LLM API usage and costs	LLM costs are significant and need attribution

2.3 Success Metrics

ID	Metric	Target	Measurement Method
sm-1	Agent Identity Coverage	100%	SPIRE registration entries / total registered agents
sm-2	Credential Exposure	0	Security scan of agent codebases
sm-3	Audit Completeness	100%	Audit log coverage analysis
sm-4	Mean Time to Revoke	< 1 minute	Revocation latency metrics
sm-5	LLM Cost Attribution	100%	Cost tracking reconciliation

---

3. Personas

3.1 End-User (Primary)

Attribute	Description
Role	Individual User
Description	Individuals who authorize agents and MCP servers to act on their behalf via ID-JAG delegation

Attribute	Description
Technical Proficiency	Medium

Goals:

- Find and use trusted AI agents safely
- Grant limited, time-bound access to personal accounts
- Maintain visibility into agent actions
- Quickly revoke access when needed

Pain Points:

- Cannot verify if an agent is trustworthy
- No control over what agents access
- Cannot see what agents did on their behalf
- Difficult to revoke access across multiple services

3.2 Agent Creator

Attribute	Description
Role	AI Agent Developer
Description	Developers who build and deploy AI agents; submit agents for KYA attestation
Technical Proficiency	Expert

Goals:

- Register agents with proper identity
- Obtain trust attestations for agents
- Access user data securely on their behalf
- Monitor agent usage and performance

Pain Points:

- Complex credential management
- No standard identity framework for agents
- Difficult to prove agent trustworthiness
- Limited visibility into agent usage

3.3 MCP Server Creator

Attribute	Description
Role	MCP Server Developer

Attribute	Description
Description	Developers who build MCP servers that connect to external services; submit MCP servers for KYA attestation
Technical Proficiency	Expert

Goals:

- Register MCP servers with proper identity
- Declare external service connections
- Obtain trust attestations
- Control which agents can invoke the MCP server

Pain Points:

- Managing credentials for external services
- No standard identity framework
- Complex agent authorization logic

3.4 Platform Administrator

Attribute	Description
Role	IT Administrator
Description	Teams managing the agent governance infrastructure; review KYA attestation submissions
Technical Proficiency	High

Goals:

- Manage organization-wide policies
- Control LLM model access and costs
- Monitor agent activity
- Respond to security incidents

Pain Points:

- No centralized agent governance
- Uncontrolled LLM costs
- Limited audit visibility
- Slow incident response

3.5 Security/Compliance

Attribute	Description
Role	Security Engineer
Description	Teams requiring audit trails and policy enforcement; manage KYA attestation policies
Technical Proficiency	High

Goals:

- Ensure regulatory compliance (SOC 2, GDPR)
- Maintain comprehensive audit trails
- Enforce security policies
- Respond to security incidents

Pain Points:

- Insufficient audit logging
- Credential sprawl across agents
- Compliance gaps
- Slow incident investigation

---

4. User Stories

4.1 End-User Stories

ID	Story	Priority	Phase
us-eu-1	As an End-User, I want to browse a catalog of verified AI agents so that I can find agents that meet my needs	High	Phase 1
us-eu-2	As an End-User, I want to view an agent's KYA attestation before granting access so that I can make informed trust decisions	High	Phase 1

ID	Story	Priority	Phase
us-eu-3	As an End-User, I want to grant an agent access to my Google Drive with read-only scope so that the agent can search my documents without modifying them	High	Phase 1
us-eu-4	As an End-User, I want to set a 24-hour expiry on agent access so that my data isn't exposed indefinitely	High	Phase 1
us-eu-5	As an End-User, I want to revoke an agent's access immediately so that I can respond to suspicious behavior	High	Phase 1
us-eu-6	As an End-User, I want to view what actions an agent performed on my behalf so that I can audit agent behavior	Medium	Phase 2

4.2 Agent Creator Stories

ID	Story	Priority	Phase
us-ac-1	As an Agent Creator, I want to register my agent with its SPIFFE ID so that it can be discovered and governed	High	Phase 1
us-ac-2	As an Agent Creator, I want to submit my agent for KYA review so that it can receive a trust attestation	High	Phase 1

4.3 Administrator Stories

ID	Story	Priority	Phase
us-ad-1	As an Admin, I want to configure which LLM models each agent can use so that I can control costs and capabilities	High	Phase 1
us-ad-2	As an Admin, I want to set monthly LLM budget limits per agent so that we don't exceed budget	Medium	Phase 2
us-ad-3	As an Admin, I want to revoke all delegations for a compromised agent so that I can respond to incidents quickly	High	Phase 1

5. Functional Requirements

5.1 Portal - End-User

ID	Title	Description	Priority	Phase
FR-P-EU-1	Agent Catalog with Trust Ratings	View catalog of available agents with trust ratings	Must	Phase 1
FR-P-EU-2	KYA Attestation Viewer	View agent KYA attestation details before granting access	Must	Phase 1
FR-P-EU-3	Scoped Delegation Grant	Grant delegation to agent with scope selection	Must	Phase 1
FR-P-EU-4	Delegation Expiry	Set delegation expiry for time-limited access	Must	Phase 1
FR-P-EU-5	Immediate Delegation Revocation	Revoke agent delegation immediately	Must	Phase 1
FR-P-EU-7	OAuth Account Connection	Connect external accounts (Google, Salesforce, etc.) via OAuth	Must	Phase 1

5.2 Portal - Agent Creator

ID	Title	Description	Priority	Phase
FR-P-AC-1	Agent Registration	Register new agent with metadata	Must	Phase 1
FR-P-AC-4	KYA Submission	Submit agent for KYA review	Must	Phase 1

5.3 KYA Attestation

ID	Title	Description	Priority	Phase
FR-KYA-2	Capability Verification	Verify agent capabilities against declared capabilities	Must	Phase 1
FR-KYA-4	Attestation Signing	Sign attestation with organizational key	Must	Phase 1

ID	Title	Description	Priority	Phase
FR-KYA-6	Attestation Revocation	Revoke attestation immediately	Must	Phase 1

5.4 Secrets Vault

ID	Title	Description	Priority	Phase
FR-SV1-1	OAuth Token Storage	Store OAuth access and refresh tokens per user per service	Must	Phase 1
FR-SV1-2	HSM-Backed Encryption	Encrypt tokens with HSM-backed keys	Must	Phase 1
FR-SV1-3	Automatic Token Refresh	Automatically refresh tokens before expiry	Must	Phase 1

5.5 SPIRE Server

ID	Title	Description	Priority	Phase
FR-SP-1	SVID Issuance	Issue SVIDs to registered agent and MCP server workloads	Must	Phase 1
FR-SP-4	Automatic SVID Rotation	Automatic SVID rotation with 1 hour default TTL	Must	Phase 1

5.6 Governance Proxy

ID	Title	Description	Priority	Phase
FR-GP-1	SPIFFE Identity Validation	Validate SPIFFE identity via mTLS for agents and MCP servers	Must	Phase 1
FR-GP-2	KYA Attestation Validation	Validate KYA attestation for all requests	Must	Phase 1

ID	Title	Description	Priority	Phase
FR-GP-3	Delegation Validation	Validate user delegation for acting-as requests	Must	Phase 1
FR-GP-4	OAuth Token Injection	Retrieve and inject OAuth tokens for SaaS APIs	Must	Phase 1
FR-GP-5	LLM API Key Injection	Retrieve and inject LLM API keys	Must	Phase 1
FR-GP-7	Model-Level LLM Policies	Enforce model-level LLM policies	Must	Phase 1
FR-GP-8	SSE Streaming Support	Support SSE streaming for LLM responses	Must	Phase 1
FR-GP-10	Request Logging	Log all requests with full context	Must	Phase 1
FR-GP-11	Credential Stripping	Strip credentials from responses	Must	Phase 1

---

6. Non-Functional Requirements

6.1 Security

ID	Title	Target	Priority	Phase
NFR-SEC-1	mTLS Inter-Service Communication	100% coverage	Must	Phase 1
NFR-SEC-2	Secrets Encryption at Rest	100% of secrets encrypted (AES-256)	Must	Phase 1
NFR-SEC-4	No Credentials in Logs	0 occurrences	Must	Phase 1
NFR-SEC-6	SOC 2 Type II Compliance	Certified	Could	Phase 3

6.2 Performance

ID	Title	Target	Priority	Phase
NFR-PERF-1	Proxy Latency	< 50ms p99 (excluding upstream)	Must	Phase 1
NFR-PERF-3	SSE Streaming Latency	< 10ms added per chunk	Must	Phase 1
NFR-PERF-4	Token Refresh Latency	< 1 second	Must	Phase 1

6.3 Scalability

ID	Title	Target	Priority	Phase
NFR-PERF-2	Concurrent Agent Connections	10,000 concurrent	Should	Phase 2
NFR-SCALE-1	Horizontal Proxy Scaling	Auto-scale based on load	Must	Phase 1
NFR-SCALE-3	Registered Agents Capacity	10,000+ agents	Should	Phase 2

6.4 Availability

ID	Title	Target	SLO	Priority	Phase
NFR-AVAIL-1	Governance Proxy Availability	99.9%	43.2 min/month error budget	Must	Phase 1
NFR-AVAIL-2	SPIRE Server Availability	99.9%	-	Must	Phase 1
NFR-AVAIL-4	Multi-Region Deployment	3+ regions	-	Could	Phase 3

6.5 Multi-Tenancy

ID	Title	Target	Priority	Phase
NFR-MT-1	Tenant Data Isolation	100% isolation test pass rate	Must	Phase 1
NFR-MT-2	Per-Tenant Encryption Keys	100% per-tenant keys	Must	Phase 1

Multi-Tenancy Specification:

Aspect	Configuration
Isolation Model	Bridge

Aspect	Configuration
Data Segregation	Schema per tenant
Encryption Model	Tenant-specific keys
Network Isolation	Shared
Noisy Neighbor Protection	Rate limiting per org: 10,000 req/min

6.6 Observability

ID	Title	Target	Priority	Phase
NFR-OBS-1	Audit Log Completeness	100% coverage	Must	Phase 1
NFR-OBS-2	Audit Log Immutability	0 tampering incidents	Should	Phase 2

6.7 Disaster Recovery

ID	Title	Target	Priority	Phase
NFR-DR-1	Governance Proxy RTO	5 minutes	Must	Phase 1
NFR-DR-2	Secrets Vault RPO	1 minute	Must	Phase 1

---

7. Roadmap

Phase 1: Foundation

Type: Generic

Goals:

- Establish core identity infrastructure with SPIFFE/SPIRE
- Implement basic KYA attestation workflow
- Deploy Governance Proxy with credential injection
- Launch Portal for all persona types
- Achieve 99.9% availability for critical services

Deliverables:

ID	Title	Type	Status
d-1-1	SPIRE Server Deployment	Infrastructure	Not Started
d-1-2	Governance Proxy MVP	Feature	Not Started
d-1-3	Secrets Vault	Feature	Not Started
d-1-4	Portal - End-User Features	Feature	Not Started
d-1-5	Portal - Agent Creator Features	Feature	Not Started

ID	Title	Type	Status
d-1-6	KYA Attestation Service	Feature	Not Started

Success Criteria:

- 100% of registered agents have SPIFFE IDs
- Proxy latency < 50ms p99
- 99.9% availability achieved
- All personas can complete core workflows

Phase 2: Scale & Compliance

Type: Generic

Dependencies: Phase 1

Goals:

- Scale to 10,000+ concurrent agents
- Implement LLM budget management
- Add advanced audit and analytics
- Prepare for SOC 2 Type I certification
- Support 100+ organizations

Deliverables:

ID	Title	Type	Status
d-2-1	LLM Budget Management	Feature	Not Started
d-2-2	Usage Analytics Dashboard	Feature	Not Started
d-2-3	Rate Limiting	Feature	Not Started
d-2-4	Audit Log Export	Integration	Not Started
d-2-5	SOC 2 Type I Preparation	Milestone	Not Started

Success Criteria:

- Support 10,000 concurrent agent connections
- LLM costs 100% attributed
- SOC 2 Type I audit ready
- 100+ organizations onboarded

Phase 3: Enterprise Ready

Type: Generic

Dependencies: Phase 2

Goals:

- Achieve SOC 2 Type II certification
- Deploy multi-region active-active
- Support hybrid deployment model
- Achieve 99.999% availability (Enterprise tier)
- Enable self-hosted deployments

Deliverables:

ID	Title	Type	Status
d-3-1	Multi-Region Deployment	Infrastructure	Not Started
d-3-2	Hybrid Deployment Model	Feature	Not Started
d-3-3	Self-Hosted Package	Feature	Not Started
d-3-4	SOC 2 Type II Certification	Milestone	Not Started
d-3-5	ISO 27001 Preparation	Milestone	Not Started

Success Criteria:

- SOC 2 Type II certified
- 99.999% availability for Enterprise tier
- Hybrid deployments operational
- Self-hosted customers successfully deployed

---

8. Technical Architecture

8.1 Overview

The architecture follows the industry-standard separation of control plane and data plane. The Control Plane handles configuration, policy, and identity issuance (SPIRE Server, Policy Store, Credential Vault). The Data Plane handles request processing and policy enforcement (Governance Gateway, SPIRE Agent). This pattern follows Istio, Kubernetes, and Consul architectures.

8.2 Integration Points

ID	Name	Type	Description	Auth Method
int-1	OAuth Providers	OAuth 2.0	User account connection (Google, Salesforce, GitHub, Microsoft 365)	OAuth 2.0
int-2	LLM Providers	REST API	Proxied LLM access (Anthropic, OpenAI, Google AI, Azure OpenAI)	API Key
int-3	Tool APIs	REST API	Proxied tool access (Serper, SerpAPI, Firecrawl)	API Key
int-4	SIEM Systems	Webhook/API	Audit log export	API Key/OAuth
int-5	SSO Providers	SAML/OIDC	Admin authentication	SAML/OIDC

8.3 Technology Stack

Layer	Technologies
Infrastructure	Kubernetes, SPIRE, HSM
Backend	Go, gRPC
Database	PostgreSQL (row-level security), Redis (caching)
Monitoring	Prometheus, Grafana, OpenTelemetry

---

9. Assumptions and Constraints

9.1 Assumptions

ID	Assumption	Risk if Invalid
a-1	Organizations will adopt SPIFFE/SPIRE as the agent identity standard	Competing standards emerge
a-2	LLM providers will maintain stable APIs	Breaking API changes require proxy updates
a-3	Enterprise customers require compliance certifications	Certification delays impact sales

9.2 Constraints

ID	Type	Constraint	Impact	Mitigation
c-1	Regulatory	Must comply with GDPR for EU users	Data residency requirements, right to deletion	EU region deployment, PII anonymization
c-2	Technical	SPIFFE/SPIRE dependency for identity	Tied to SPIFFE ecosystem evolution	Active SPIFFE community participation
c-3	Regulatory	SOC 2 Type II requires 6+ months of evidence	Cannot certify immediately	Start evidence collection early in Phase 1

9.3 Dependencies

ID	Name	Type	Status
dep-1	SPIFFE/SPIRE Project	External	Available
dep-2	LLM Provider APIs	External	Available
dep-3	HSM Provider	Vendor	Pending

---

10. Out of Scope

- Agent cognition/reasoning (handled by agent frameworks like LangChain, AutoGen)
- LLM model training or fine-tuning
- End-user facing chatbot UI
- Data labeling or annotation tools
- Agent marketplace/app store (future consideration)

---

11. Risk Assessment

ID	Risk	Probability	Impact	Mitigation	Status
r-tm-1	Stolen SVID enabling agent impersonation	Low	High	Short SVID TTL (1hr), workload attestation	Mitigated
r-tm-2	Compromised agent binary performing unauthorized actions	Medium	High	KYA code review, container signing	Open
r-tm-3	Delegation scope escalation granting excess permissions	Low	Medium	Strict scope validation, UI confirmation	Mitigated
r-tm-5	Secrets vault breach exposing credentials	Low	Critical	HSM backing, encryption at rest, access logging	Mitigated
r-tm-7	Governance Proxy bypass enabling direct API access	Low	High	Network policies, no credentials in agents	Mitigated

ID	Risk	Probability	Impact	Mitigation	Status
r-tm-8	Insider admin abuse leading to data exfiltration	Low	High	Audit logging, separation of duties, MFA	Open
r-market-1	Competing standard emerges for agent identity	Medium	High	Active participation in SPIFFE community, extensible architecture	Open

---

12. Market Opportunity

12.1 TAM/SAM/SOM

Metric	2030 Estimate
TAM (Agent Identity & Governance)	$9-10 billion
SAM (Enterprise Agentic AI Infrastructure)	$3.7 billion
SOM (Year 5)	$250 million

12.2 CAGR

Market	CAGR
AI Agents	46.3%
Agentic AI	43.8%
AI Governance	49-51%

12.3 Market Gap

No existing solution provides end-to-end agent identity (SPIFFE), attestation (KYA), credential injection, and observability in a unified control plane.

---

13. Deployment Options

13.1 Deployment Models

Model	Control Plane	Gateway	SLA	Target Customer
Managed SaaS	Vendor	Vendor	Up to 99.999%	Most customers
Hybrid	Vendor	Vendor-in-Customer-VPC	99.99%	Compliance-sensitive
Self-Hosted	Customer	Customer	Customer's SLA	Air-gapped, regulated

13.2 Pricing Tiers

Tier	Monthly Price	SLA
Starter	$500	99.9%
Business	$2,000	99.95%
Enterprise	$8,000	99.99%
Premium	$25,000	99.999%

---

14. SLA Definitions

14.1 Service Level Objectives

Service	Metric	Business Tier	Enterprise Tier
Governance Proxy	Availability	99.99%	99.999%
Governance Proxy	Latency (p99)	< 50ms	< 25ms
SPIRE Server	SVID issuance	< 500ms	< 200ms

14.2 SLA Exclusions

- Upstream service failures (LLM providers, SaaS APIs)
- Scheduled maintenance windows (72-hour notice)
- Customer-caused issues
- Force majeure events

---

15. Glossary

Term	Definition
SPIFFE	Secure Production Identity Framework for Everyone - a set of standards for identifying and securing communications between services
SPIRE	SPIFFE Runtime Environment - the reference implementation of SPIFFE
SVID	SPIFFE Verifiable Identity Document - an X.509 certificate that encodes a SPIFFE ID
KYA	Know Your Agent - attestation framework for verifying AI agent trustworthiness, capabilities, and security posture
Delegation	User authorization for an agent to act on their behalf with specific scopes and time limits
mTLS	Mutual TLS - both client and server present certificates for authentication
SSE	Server-Sent Events - streaming protocol used for LLM responses
MCP	Model Context Protocol - protocol for connecting AI agents to external tools and services
Control Plane	Infrastructure components responsible for configuration, policy, and identity issuance
Data Plane	Infrastructure components responsible for request processing and policy enforcement

---

Generated from structured PRD JSON format