SOLUTION BRIEF

The Security Layer Between Your AI Agents and the Internet

Pipelock gives security teams an enforcement and inspection boundary for AI agent traffic across HTTP, WebSocket, and MCP.

Open source core · **Signed releases + SBOM** · **OWASP mappings** · **Proxy + Scan API**

Why Teams Are Evaluating Agent Egress Controls Now

- Agents now have credentials, tool access, and outbound network reach.
- Prompt injection and poisoned tools can turn normal workflows into exfiltration paths.
- Most teams still lack a dedicated boundary that inspects and controls what agents send out.

What Pipelock Is

Pipelock sits between the agent and external systems. It does not replace model guardrails or sandboxing. It adds a separate runtime security layer at the traffic boundary.

```
Agent (secrets, no network) → Pipelock (inspection + policy) →
Internet / MCP / APIs
```

The agent process keeps credentials but cannot reach the internet directly. Pipelock has full network access but holds no agent secrets. Enforced at the deployment layer: Docker, Kubernetes NetworkPolicy, iptables, or macOS PF.

What Pipelock Is Designed to Reduce

Credential exfiltration. 46 DLP patterns across URLs, headers, and request bodies. Decodes base64, hex, URL-encoding, and Unicode evasion before matching.

Prompt injection. 6-pass normalization pipeline handles zero-width characters, homoglyphs, leetspeak, and encoded payloads in fetched content and tool responses.

SSRF and metadata access. Blocks private IPs, link-local, cloud metadata endpoints. DNS rebinding protection.

MCP tool poisoning. Scans tool descriptions for hidden instructions. Detects mid-session changes. 17 pre-execution policy rules.

Crypto address poisoning. ETH/BTC/SOL/BNB validation against allowlist. BIP-39 seed phrase detection with checksum.

11

layer scanning pipeline

46

DLP patterns with checksum validators

17MB

single static binary

> *“Fail-closed everywhere. Timeouts, parse errors, and ambiguous state all default to block.”*

Defense in Depth

Sandbox isolates execution
Guardrails shape model output
Pipelock inspects the traffic

Framework Mapping

- OWASP Agentic AI Top 10
- OWASP LLM Top 10
- EU AI Act (Art. 9, 12-15, 26)
- NIST AI RMF 1.0

Fits Into Existing Agent Workflows

Mode	What It Does
Proxy	Outbound HTTP/HTTPS scanning via HTTPS_PROXY. Optional TLS interception.
MCP proxy	Wraps stdio or HTTP MCP servers with bidirectional scanning.
IDE hooks	One command for Claude Code, Cursor, or VS Code.
Scan API	POST /api/v1/scan for out-of-band verdicts from orchestrators, CI/CD, SIEM.

Use proxy mode in the data plane. Use Scan API when you need a verdict service in the control plane.

Built for Evaluation, Control, and Auditability

- **Fail-closed behavior** on parse errors, context cancellation, timeouts, and ambiguous state.
- **Structured audit logs** with MITRE ATT&CK technique IDs. Webhook, syslog, and OTLP emission. 30 Prometheus metric families.
- **Policy enforcement at runtime,** not just prompt-time guidance.
- **Signed artifacts** with cosign signatures, CycloneDX SBOM, and SLSA v1.0 provenance attestation.
- **Documented limitations** and honest gap analysis. Published OWASP coverage mappings.

Serious Engineering, Not a Demo

- 7,000+ tests with race detection across all proxy, scanner, MCP, and policy paths. 90%+ coverage.
- Private adversarial testing (800+ cases) against the production binary before every release
- Cross-platform release validation (Linux, macOS, Windows on amd64 + arm64)
- Open source core with public docs, threat model, and evasion resistance test matrix
- Single static binary, 17MB, 12 direct Go dependencies

Typical Evaluation Scenarios

- Teams deploying AI coding agents internally (Claude Code, Cursor, VS Code)
- Enterprises adopting MCP-based tooling across multiple teams
- Security teams reviewing agent egress risk
- Control-plane builders who need a verdict service via API

Kill Switch

4 independent sources: config file, SIGUSR1, sentinel file, remote API. Any one active blocks all traffic. OR-composed. Agent cannot self-deactivate.

Supply Chain

- Cosign-signed releases
- CycloneDX SBOM per release
- SLSA v1.0 provenance
- OpenSSF Scorecard 8.7/10
- OpenSSF Best Practices Silver
- 12 direct Go dependencies
- Apache 2.0 (auditable core)

Licensing

Community — Free forever. Full scanning engine, all DLP patterns, all proxy modes.

Pro — Multi-agent coordination with named profiles and per-agent policies.

Enterprise — Custom annual. Multi-team governance and support.

Evaluating Pipelock for your team?

We can help scope deployment mode, logging, and policy fit for your environment.

pipelab.org/enterprise · luckypipe@pipelab.org · github.com/luckyPipewrench/pipelock