Zero Trust for *Autonomous Agents*

Isolating AI Workloads on Kubernetes

About the *speaker*

Senthalan Kanagalingam

Technical Lead · WSO2

8+ years in enterprise IAM

Started with - WSO2 Identity Server & Asgardeo

THEN

AppDevIAM lead — WSO2 Developer Platform

The internal platform that evolved into OpenChoreo (CNCF Sandbox)

NOW

Core maintainer — ThunderID

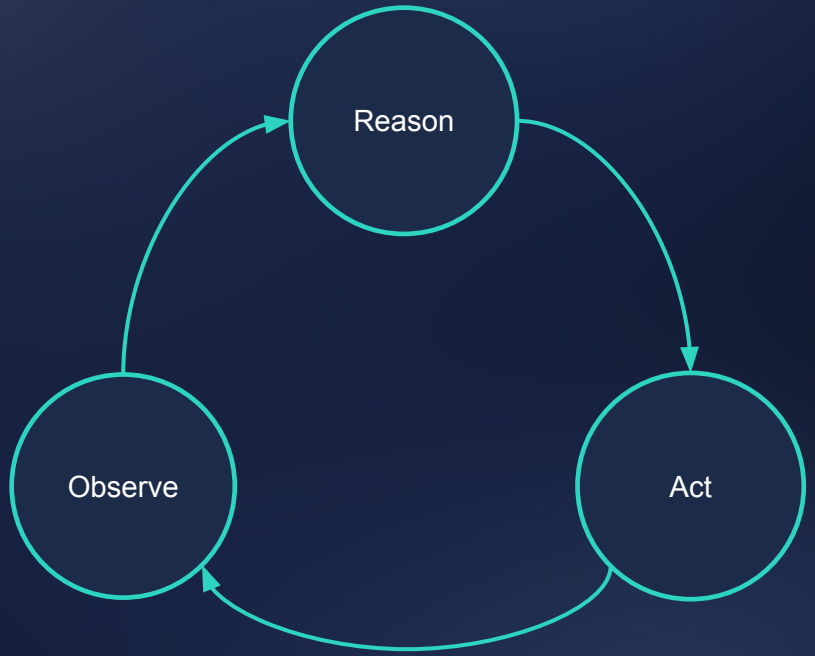
High-performance open-source identity stack

Meet your newest *coworker*

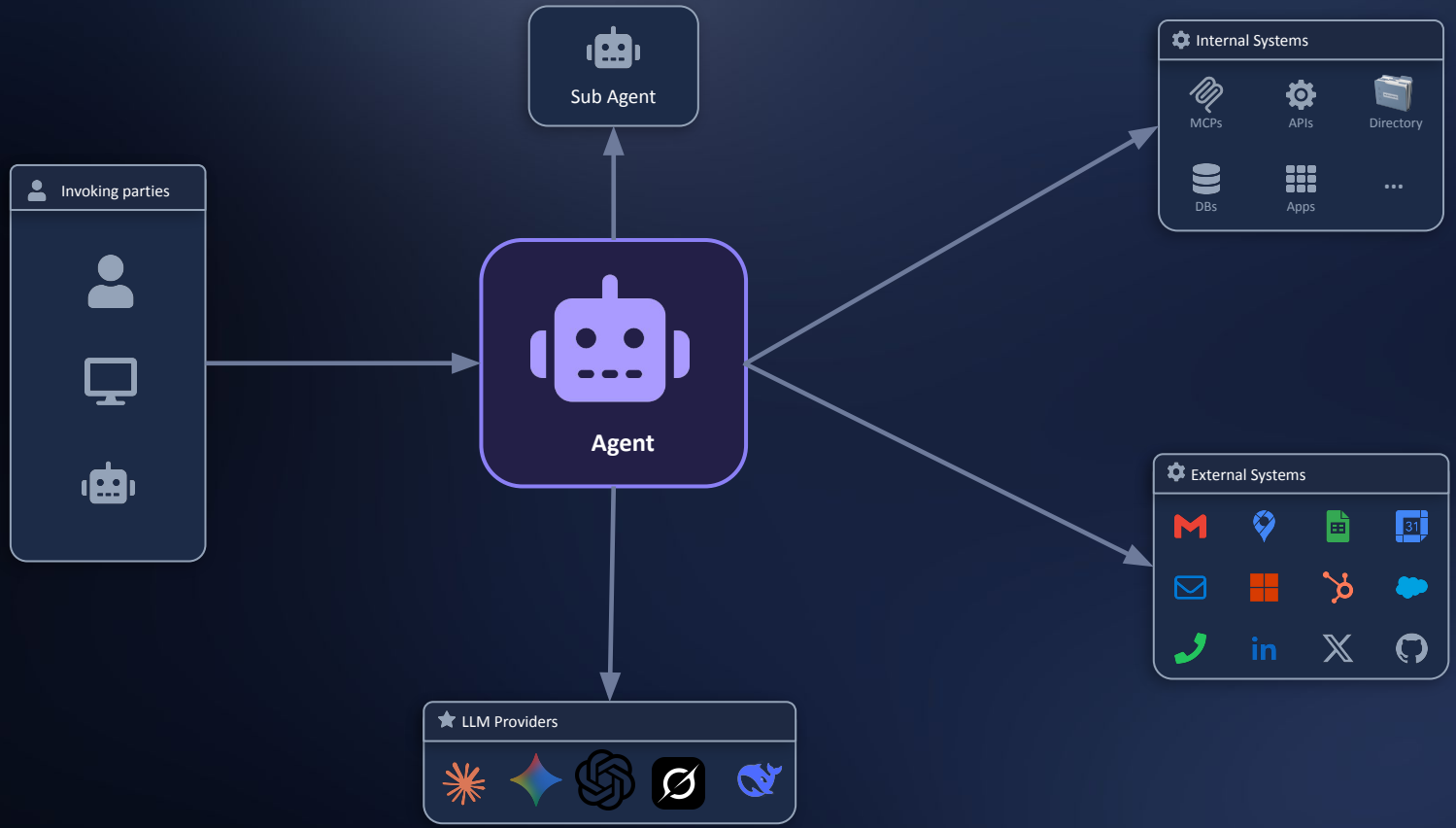
An autonomous agent

Helpful. Tireless. Self-directed

0 sleep	µs speed	∞ eager



The agent's *reach*



Zero Trust: the *lens*

Verify explicitly

Every request is authenticated and authorized on its own merits — no free pass for sitting inside the network.

Least privilege

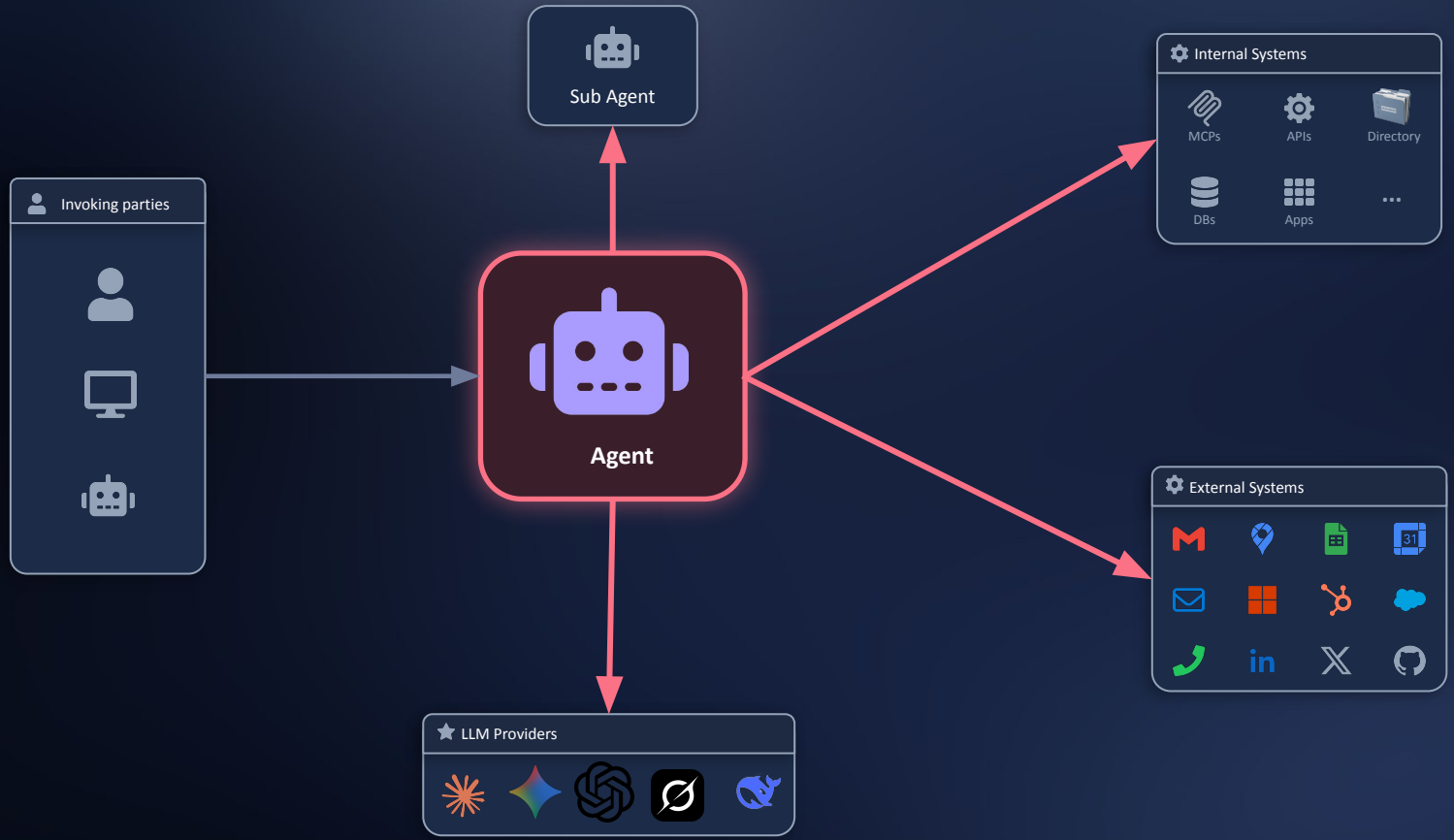
Grant the minimum access needed — nothing more, and only for as long as it's needed.

Assume breach

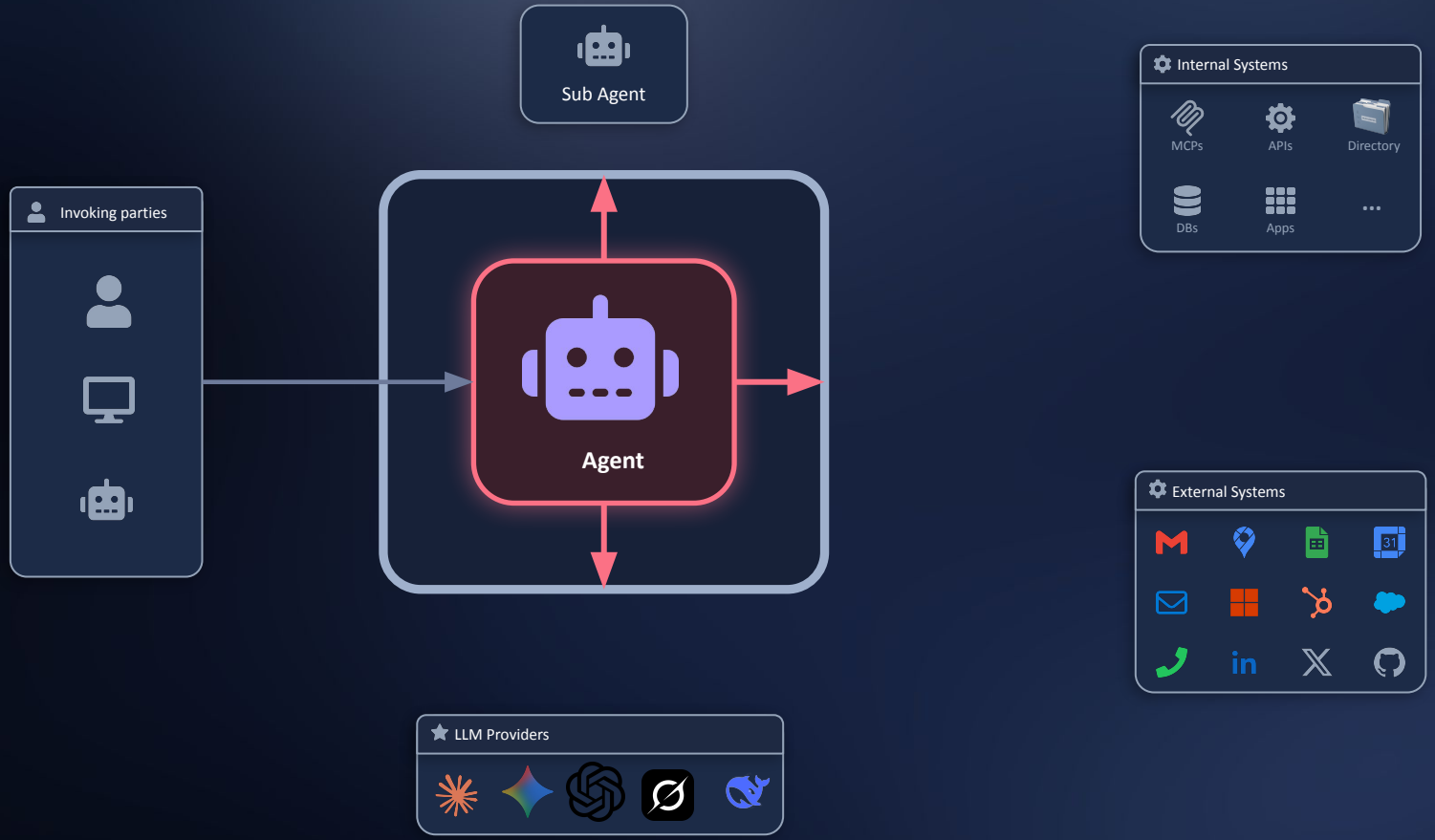
Design as if the attacker is already in. Contain what any single compromise can reach.

For agents, **assume breach** stops being a precaution and becomes a description.

How *far?*


Sub Agent
Internal Systems
MCPs
APIs
Directory
DBs
Apps
...
Invoking parties
Agent
External Systems
LLM Providers

As far as the *wall*

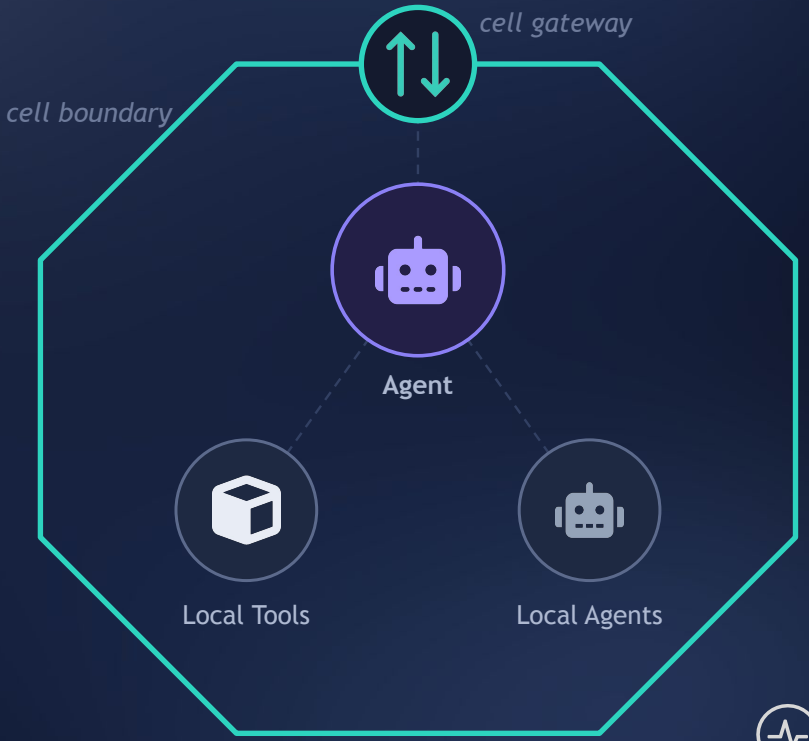


The *cell*, built around the agent

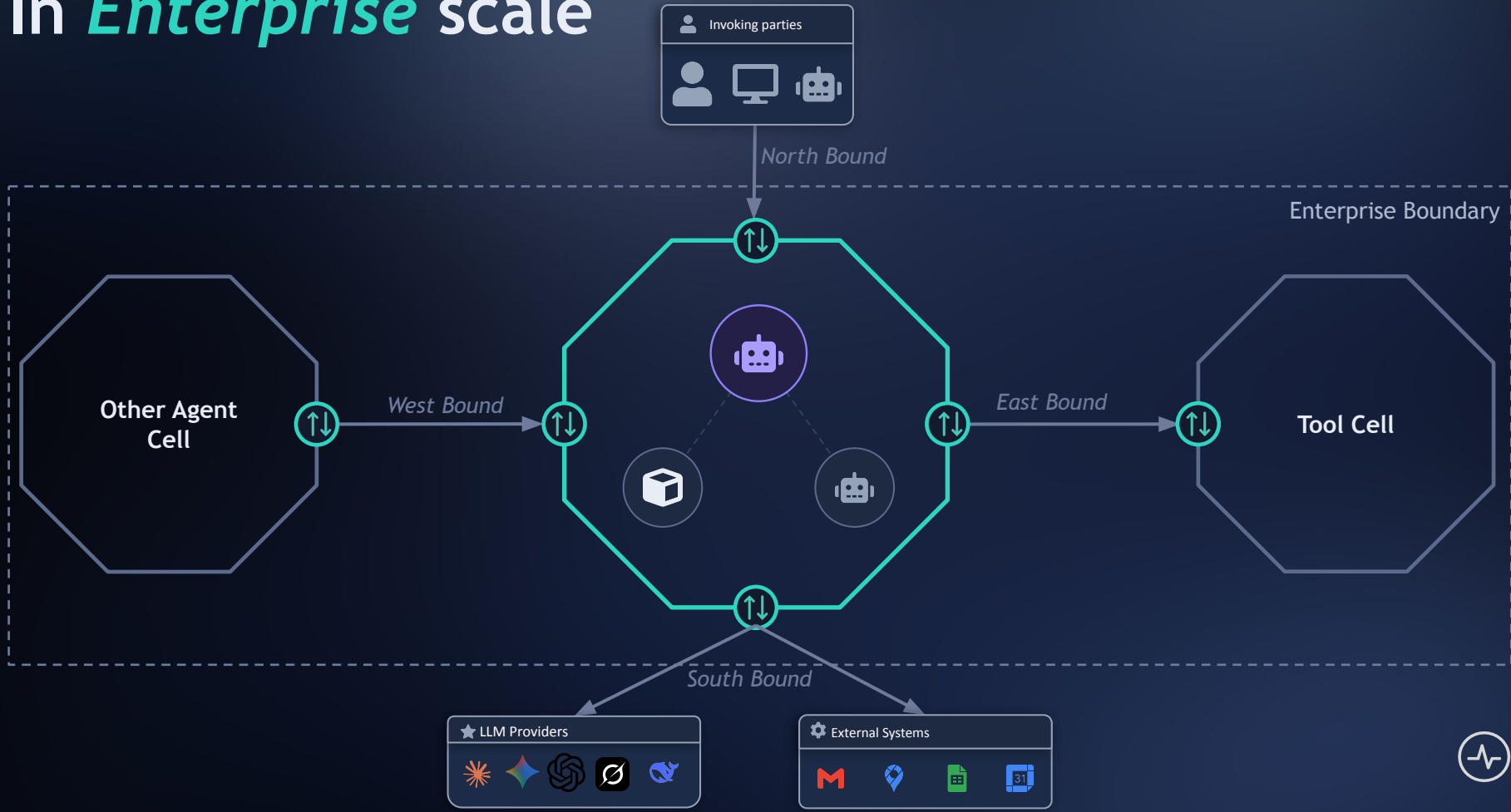
A decentralized way to group components into an independently deployable, immutable unit — the cell.

- **A hard boundary** — components are reached only through the cell, never directly
- **Gateway** — all traffic in or out passes the controlled entry point
- **Bounded by domain** — maps to one business capability (DDD), owned by one team

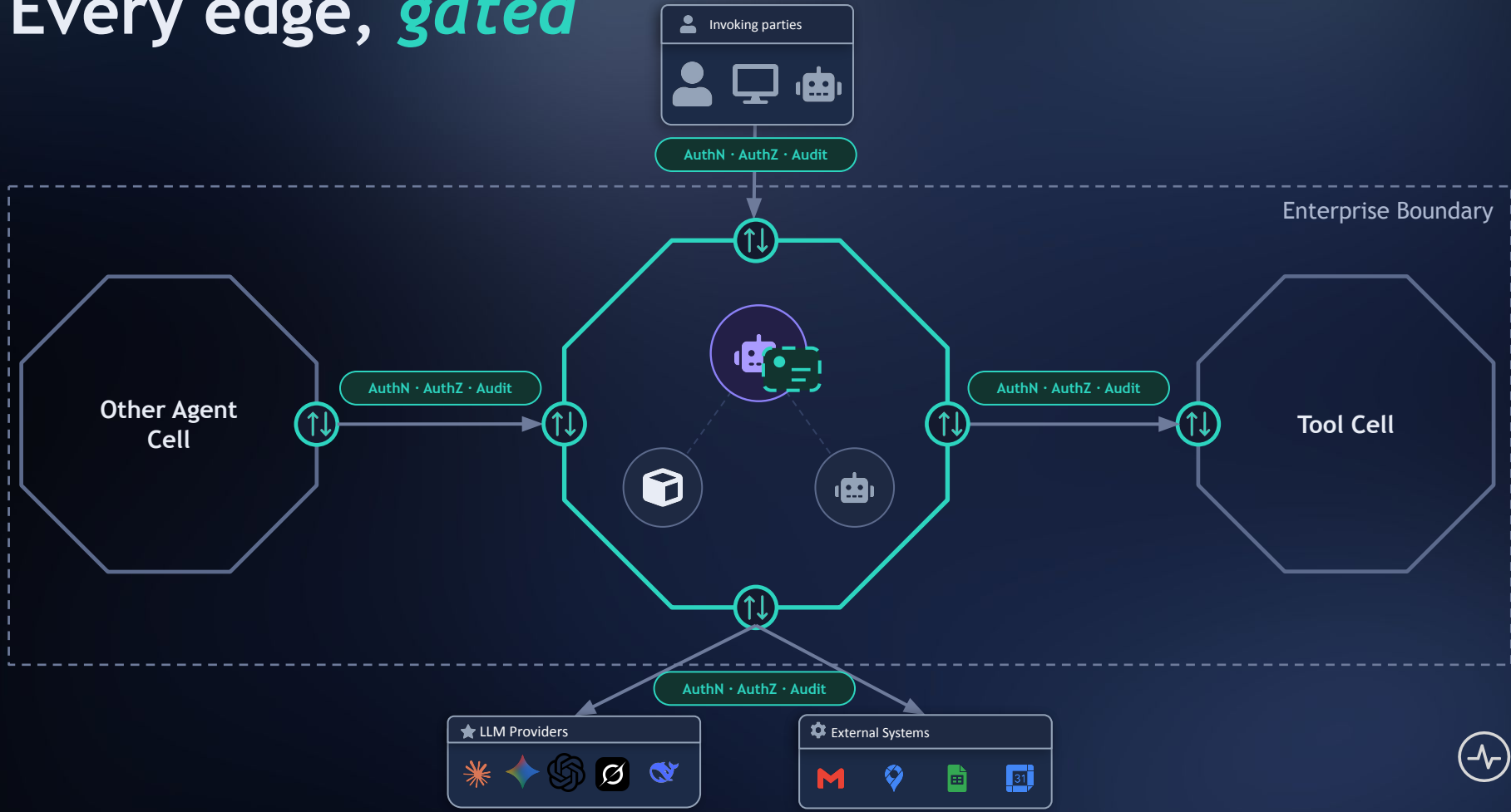
Introduced by WSO2's Asanka Abeysinghe & Paul Fremantle, 2018.



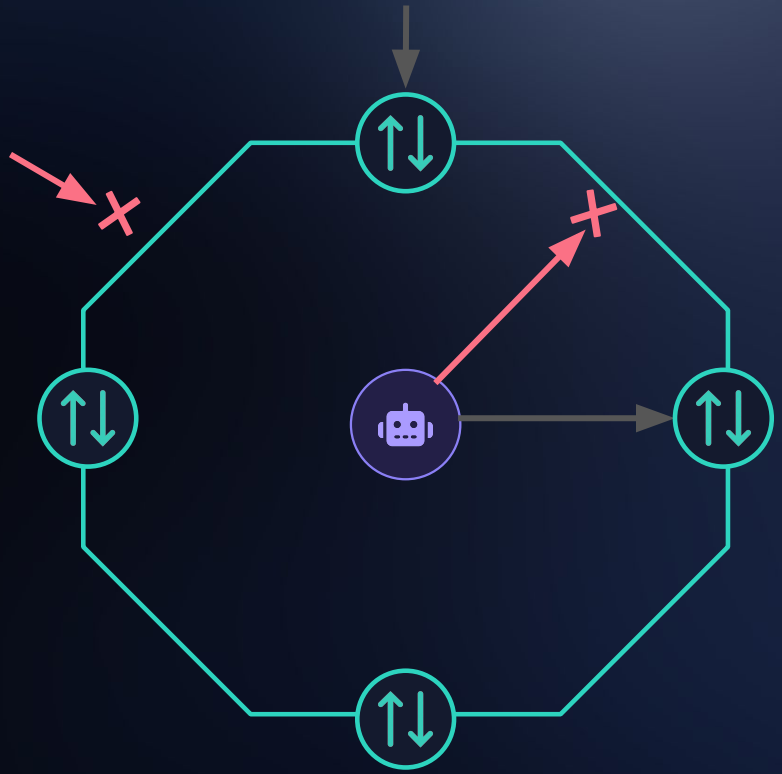
In *Enterprise* scale



Every edge, *gated*



The *cell* in Kubernetes



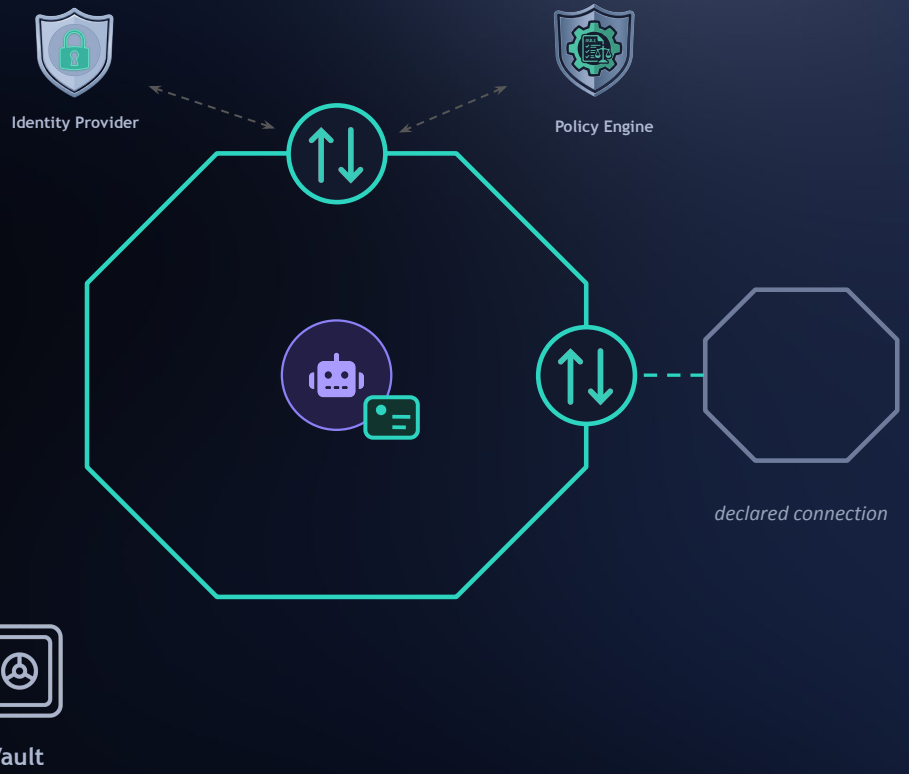
- Cell = Kubernetes namespace
- Network policy, default-deny

  cilium ANTREA Kube-OVN
- Encrypted traffic

  Istio cilium LINKERD
- Gateway as the PEP

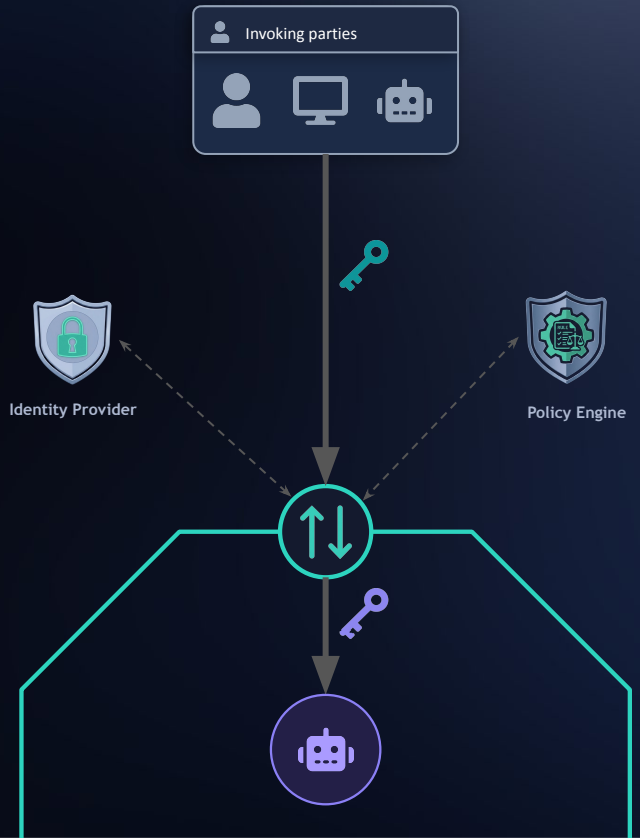
  envoy gateway CONTOUR kgateway

Administration — *declared up front*



- **Connections declared**
  allowed paths compile to network policy
- **Permission map per action**
  the gate never guesses
- **Registered in the IDP**
  owner · auth mechanism · lifecycle kill switch
- **Policies in the engine**
  Dynamic policy evaluation per request
- **Secrets stay in the vault**
  the agent holds references, never credentials

Initiator to Agent



- **Gateway trusts an IDP**
- **Every request carries a valid token**
- **Real time authorization policy check**

the policy engine decides per request

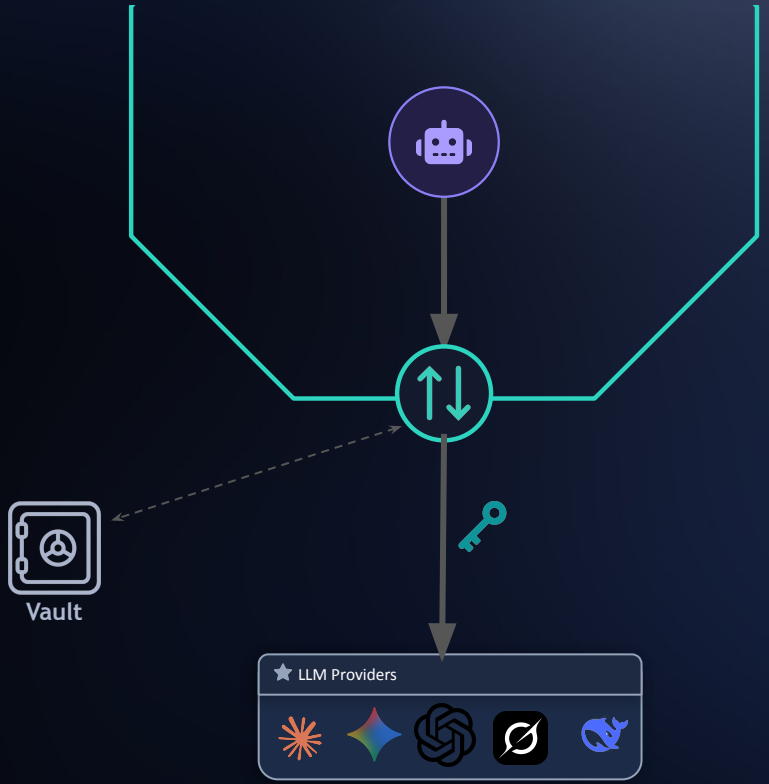
- **The token never reaches the agent**

the agent gets only a reference token

- **Gateway can do the OAuth dance**

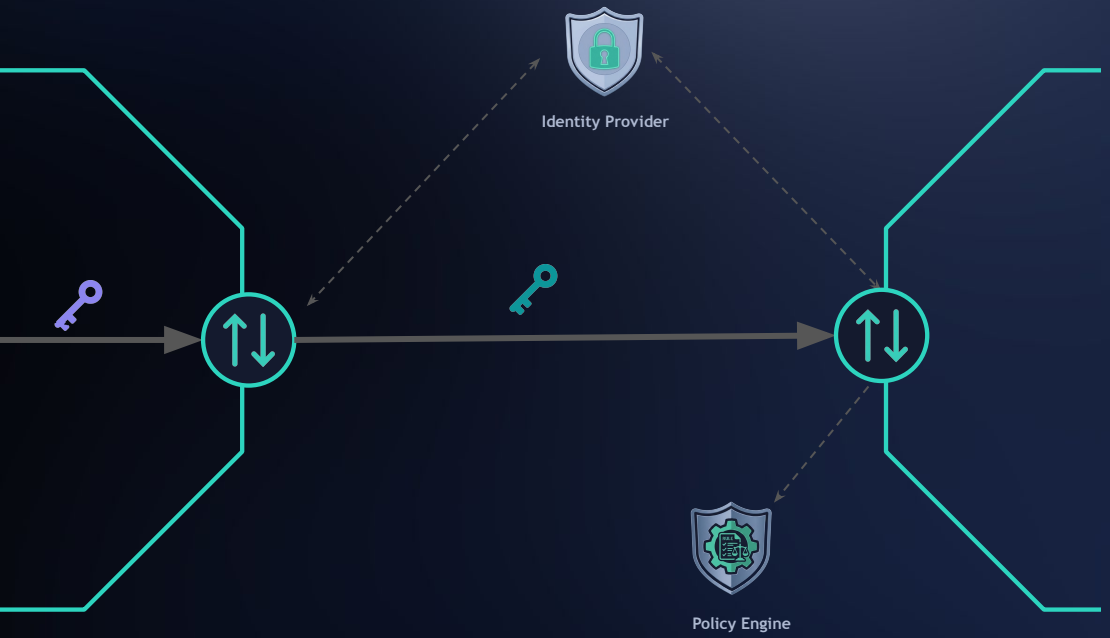
agent will be security protocol agnostic

Agent to AI Providers



- **Provider & model allow-list**
  only approved models, declared at setup
- **Vault-injected credentials**
  the agent never holds a standing API key
- **Usage limits & budgets**
- **PII redaction & guardrails**
  sensitive data filtered at the gate
- **Attributed audit trail**

Agent to Tool

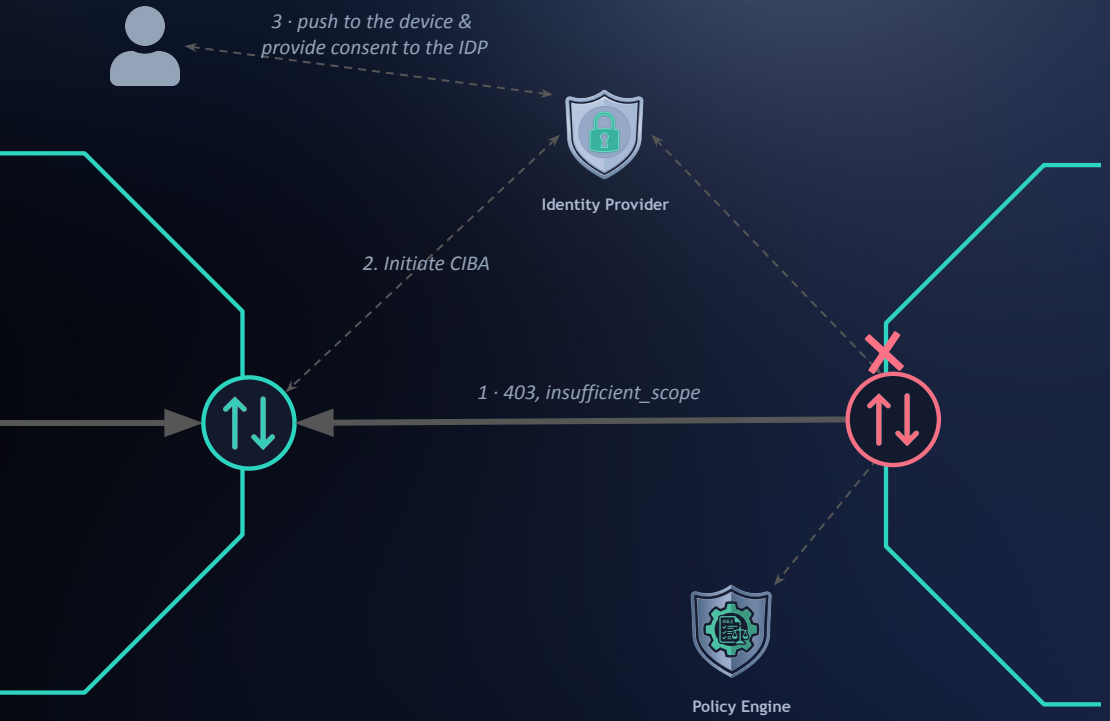


- **Allow-list first**
  not on the agent's list → deny
- **Token exchange**
  a downscoped token: sub user, act agent
- **Intersection authorization**
  user ⋂ agent ⋂ policy — the user is the ceiling
- **The confused deputy dies here**
  the check: does the user have this scope?
- **Authorization failure? Escalate**
  insufficient scope or insufficient consent

Human in the Loop — user present

- **The tool gate's answer**
  Need more scope or consent
- **Challenge flows back to the user**
  insufficient_scope — up the same path it came
- **Consent runs user ↔ IDP**
  the gateway drives the screen — not the agent
- **New token, retry**
  the added scope is minted; the call replays
- **The agent only hears**
  blocked, pending — then approved

Human in the Loop — user away



- **No session to step up**
  the agent is running in the background with user context
- **CIBA — backchannel consent**
  an OpenID standard for out-of-band approval
- **Push to the user's device**
  approve · deny · timeout
- **The gate polls, then decides**
  approved ⟶ new token ⟶ retry
- **High-stakes actions only**
  not every call needs a human

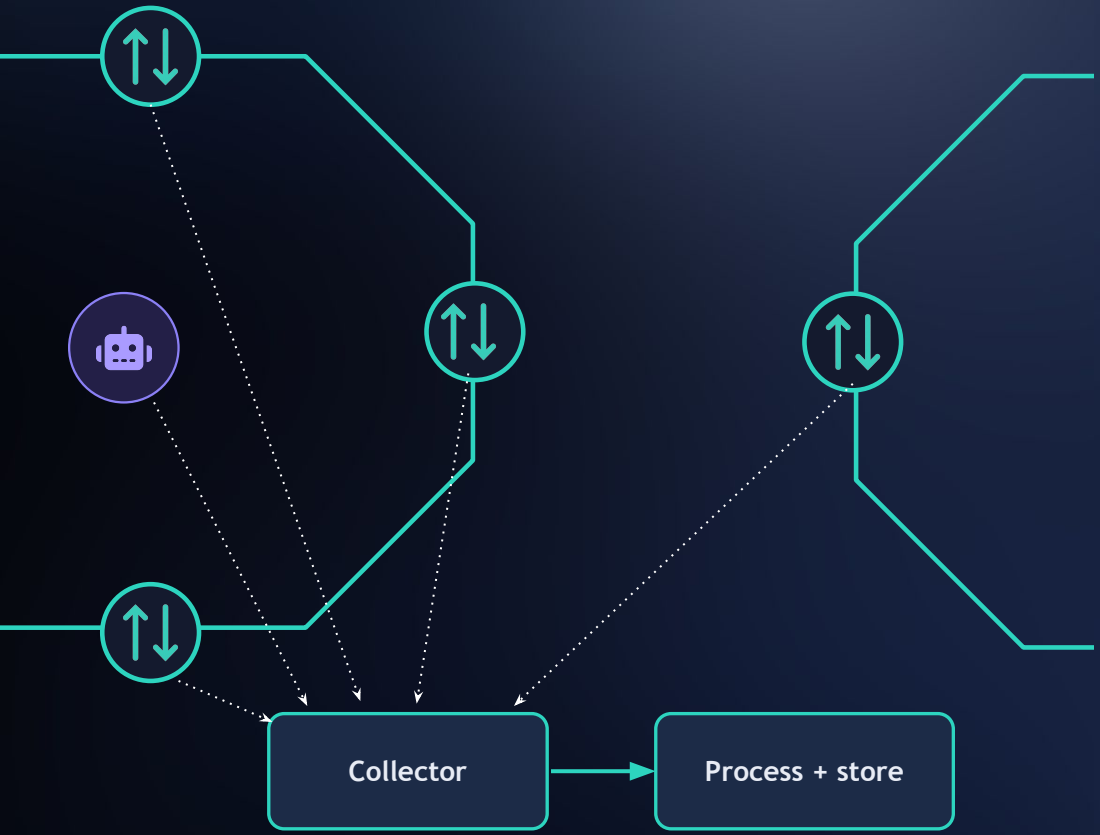
Agent to Agent

- **Mostly similar to Agent to Tool**
- **Token exchange** downscoping & act chaining

```
"sub": "alice",
"act": {
  "sub": "analytics_agent",
  "act": {
    "sub": "reporting_agent"
  }
},
"scope": "write_report"
```

- **Authority can only narrow** never widen
- **Ingress gateway performs its own check**

Audit: one cross-boundary sweep



- **Observable by default**
- **Collect & enrich**

OpenTelemetry fluentd Hubble

- **Metrics & traces**

Prometheus JAEGER Thanos

- **One trace ID joins both layers**

even denials — dropped packet, full attribution

What this pattern *buys you*

Platform engineers

security you can operate

- Policy lives at the platform — not in agent code
- Blast radius capped by default-deny
- One kill switch — suspend, and tokens die
- One trace — the full forensic chain

Developers

agent · app · MCP — security you don't write

- Never touch a token or a credential
- On-behalf-of, step-up, CIBA — from the platform
- Declare tool → action → scope, get authz
- Swap models — security isn't in the prompt

Developers build the agent — the platform builds the trust.

Two jobs, *two projects*

ISOLATION

An open-source internal developer platform for Kubernetes, built on cell-based architecture.

In this architecture: bounds what chains can exist.

Cell-based isolation
one project ⟶ one isolated namespace per environment

Default-deny, declared paths only
Cilium / eBPF policies generated from declarations

IDENTITY

An open-source, lightweight and high-performing IAM built for humans, applications — and AI agents.

In this architecture: bounds a chain's authority.

Agents as first-class identities
registered, with a liable entity accountable for each

Everything defined as YAML
agents, roles, resources — declarative, version-controlled

A packet reaches a tool only if **the network permits the path** AND **the token proves the authority.**

Demo

Questions

Thank you