AVIATRIX

TECHNICAL BRIEF

Contain Azure AI Foundry Agents

Powered by Aviatrix + Microsoft

Aviatrix Distributed Cloud Firewall · Reference Architecture for Azure AI Foundry Agents

Aviatrix Validated Containment Architectures are lab-tested containment deployment blueprints for the AI platforms enterprises are actually running. They are ship-ready, policy-included, and validated before they arrive. This Validated Containment Architecture covers Azure AI Foundry Agents.

Threat Context

Azure AI Foundry BYOVNet mode presents a challenge for security teams: it places agent workloads in a customer-managed virtual network via subnet delegation. The hosted agent subnet has a no-control egress path to the internet.

Three attack vectors are relevant to this deployment:

- Unconstrained tool-call egress: every tool call, MCP server connection, and Code Interpreter outbound request exits through the delegated subnet to any internet destination.
- Control-plane TLS conflict: applying TLS inspection broadly to the hosted agent subnet breaks Azure Active Directory (AAD) token acquisition, container image pulls, and Azure infrastructure calls. The enforcement solution must use per-rule decryption policy to optionally selectively inspect tool-call traffic and explicitly bypass control-plane destinations.
- Lateral movement to adjacent spokes: a compromised agent can attempt to reach databases, internal APIs, and other workloads in connected VNet spokes unless explicit East-West policy is enforced.

LAB-VALIDATED THREAT SCENARIO

The Aviatrix lab validated this architecture against the following scenarios: prompt injection via MCP tool response exfiltrating to an attacker-controlled HTTPS endpoint (blocked by default-deny WebGroup rule); lateral movement from the agent subnet to an adjacent spoke private endpoint (blocked by East-West SmartGroup deny); and hosted agent runtime startup with Aviatrix in-path. Each scenario maps to an OWASP LLM Top Ten entry and the DCF rule that closes it.

Prerequisites

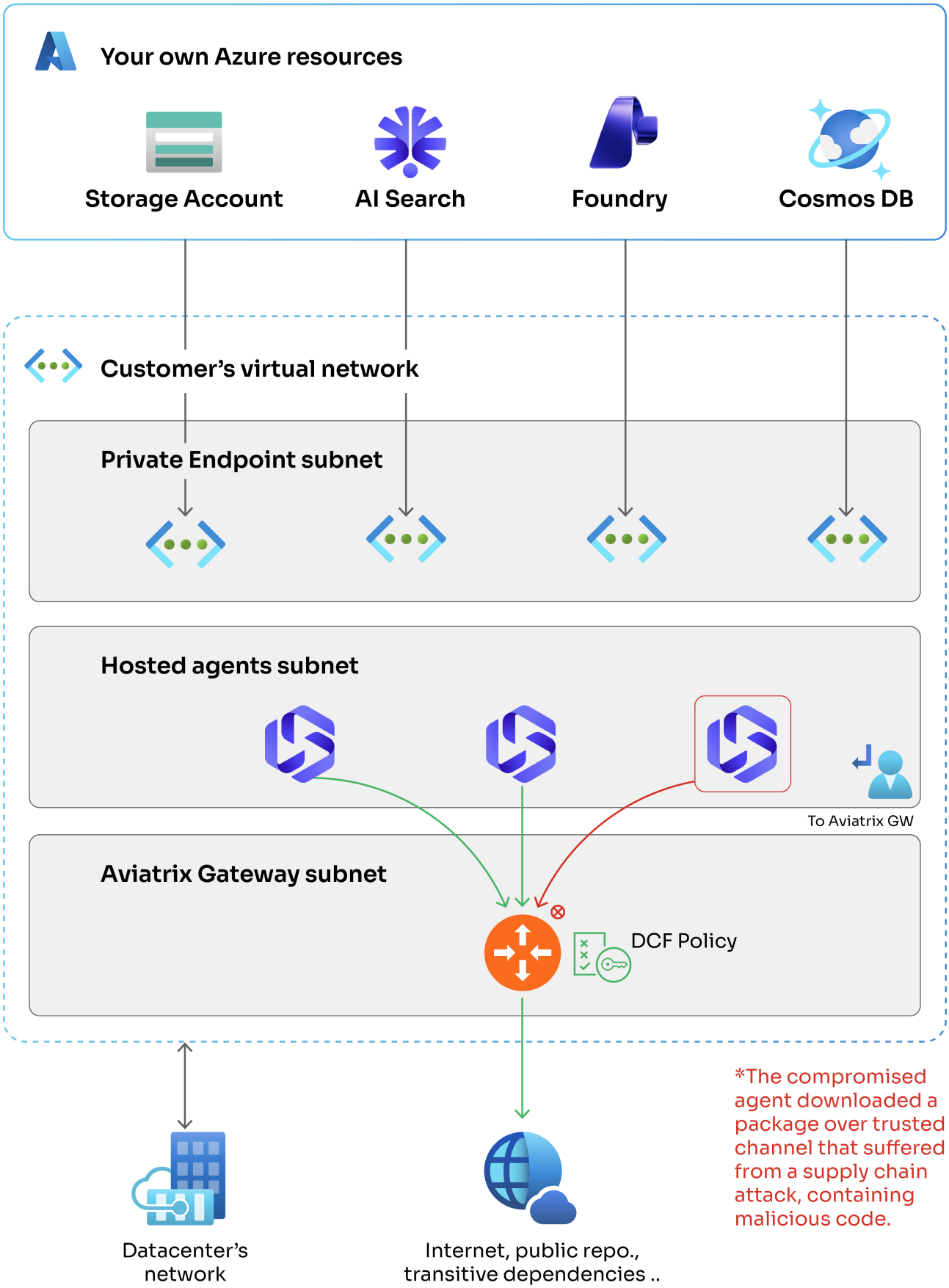
Before configuring Distributed Cloud Firewall (DCF) enforcement, verify the following are in place:

- **Aviatrix Controller 9.0 or later**
  Controller 9.0 is required for selective TLS decryption and per-rule decryption policy scope.
- **Aviatrix Spoke Gateway deployed in the Foundry spoke VNet** in the same Azure region as the Foundry account (required by Microsoft). Minimum one high availability (HA) pair, Standard_B2ms or equivalent.
- **Azure AI Foundry Hosted Agent in BYOVNet mode**
  The agent subnet must be delegated to Microsoft.App/environments.
- **User-Defined Route (UDR) on the hosted agent subnet:** 0.0.0.0/0 next-hop pointing to the Spoke Gateway private IP will be updated by Aviaitrix control plane.

What Ships on Day One

The architecture is one repository, one Terraform module, one set of validated controls. Everything required to deploy is in the box.

- **Insertion pattern**
  Architecture diagram and configuration notes showing an Aviatrix Spoke Gateway deployed in the Foundry spoke VNet, with a user-defined route on the hosted agent subnet routing all egress through the Spoke Gateway. Covers both the Transit-attached deployment for existing Aviatrix customers and the standalone Spoke Gateway deployment for organizations with no Transit in place.
- **SmartGroup model**
  SmartGroups keyed to Foundry's identity primitives: a source SmartGroup for the foundry-agents subnet, Service Tag-based SmartGroups for hosted agent platform traffic, and a destination SmartGroup for approved MCP server FQDNs and sanctioned tool-call APIs.
- **Baseline Distributed Cloud Firewall policy pack**
  Six prioritized DCF rules enforcing default-deny egress: threat intelligence deny at highest priority; ACA platform traffic permitted via FQDN WebGroup and Service Tag SmartGroups ; approved tool-call destinations permitted with optional payload-level inspection; catch-all FQDN deny; and East-West deny blocking lateral movement from the agent subnet to adjacent workload spokes if connected to a transit.
- **Bill of materials**
  Aviatrix Spoke Gateway (one per Foundry VNet, same Azure region as the Foundry account required by Microsoft), UDR on the delegated agent subnet with Aviatrix programming 0.0.0.0/0 next-hop to the Spoke Gateway private IP, and Aviatrix Enterprise Transit for customers connecting to an existing fabric and requiring East/West traffic filtering. Validated with Controller 8.2+.
- **GitHub repository**
  Full Terraform for the complete deployment, including the WebGroup allowlist, DCF policy definitions, SmartGroup configurations. Policy-as-code structured to live in the same repository as the agent configuration and deploy in the same pipeline run.



Architecture: Azure AI Foundry BYOVNet with Aviatrix Spoke Gateway in the egress path.

SmartGroup and WebGroup Design

The policy model uses five SmartGroup objects and two WebGroup objects. All objects are defined in the Aviatrix Controller and referenced by name in the DCF policy pack.

Object	Type / Scope / Purpose
foundry-agents	Subnet SmartGroup – matches the hosted agent subnet CIDR in the Foundry spoke VNet. Source identity for all egress rules.
aca-requirements-fqdns	WebGroup – Hosted agent platform FQDNs required for runtime operation. Includes, non limited to: mcr.microsoft.com, *.data.mcr.microsoft.com, packages.aks.azure.com, acs-mirror.azureedge.net, *.identity.azure.net, login.microsoftonline.com, *.login.microsoftonline.com, *.login.microsoft.com, login.microsoft.com, *.blob.core.windows.net. Applied with no decryption because of Microsoft certificate pinning.
other-spoke-resources	SmartGroup – matches private endpoints and resources in adjacent workload spokes connected via Enterprise Transit. Destination identity for the East-West deny rule.
default-threatgroup	Aviatrix-managed threat intelligence feed – known malicious destinations, command-and-control infrastructure, newly registered domains. Applied as highest-priority deny rule.

Distributed Cloud Firewall Policy Pack

Six rules in priority order. First match wins. Every rule is logging-enabled and writes to CoPilot Distributed Cloud Firewall logs. Deploy all rules in monitor (watch) mode first; promote to enforcement rule by rule after validating against production traffic.

Pri	Rule Name	Source	Destination	Action / TLS / Notes
01	deny-threat-intel	foundry-agents	default-threatgroup	**DENY · – · Highest priority; blocks known malicious destinations, C2 infrastructure, newly registered domains before any permit rule evaluates**
02	allow-aca-fqdn	foundry-agents	aca-requirements-fqdns	**PERMIT · DECRYPT_NOT_ALLOWED · hosted agent platform FQDNs required for runtime. Must precede default-deny.**
03	allow-aca-svctag	foundry-agents	aca-requirements-svctag SmartGroups	**PERMIT · DECRYPT_NOT_ALLOWED · Azure Service Tags for MCR, AzureFrontDoor, ACR, AAD. Auto-updates with Microsoft IP changes.**

04	allow-tool-calls	foundry-agents	foundry-tool-calls WebGroup	**PERMIT · DECRYPT_ALLOWED and optional · Approved MCP servers, sanctioned APIs, Code Interpreter destinations.**
05	default-deny-internet	foundry-agents	Any unlisted FQDN / IP	**DENY · – · All non-private-endpoint egress not matched above. Every deny logged to CoPilot FlowIQ with rule name and destination.**
06	deny-east-west	foundry-agents	other-spoke-resources	**DENY · – · Lateral movement from Foundry agent subnet to adjacent spoke private endpoints and workloads.**

ACA Platform WebGroup — Detailed FQDN List

The aca-requirements-fqdns WebGroup must contain these entries, sourced from the Microsoft Azure Container Apps firewall rules documentation. All entries use DECRYPT_NOT_ALLOWED. Substitute [region] with the Azure region identifier (for example, eastus).

Scenario	FQDNs	TLS Handling
MCR (all scenarios)	mcr.microsoft.com, *.data.mcr.microsoft.com	DECRYPT_NOT_ALLOWED
AKS infrastructure (all scenarios)	packages.aks.azure.com, acs-mirror.azureedge.net	DECRYPT_NOT_ALLOWED
Managed identity	*.identity.azure.net, login.microsoftonline.com, *.login.microsoftonline.com, *.login.microsoft.com	DECRYPT_NOT_ALLOWED
ACR (Hosted Agents)	login.microsoft.com, *.blob.core.windows.net <- optional as per Microsoft, not in the default WebGroup.	DECRYPT_NOT_ALLOWED

Azure Service Tag SmartGroups

Create one SmartGroup per Service Tag using the Azure IP external match pattern. These SmartGroups auto-update as Microsoft updates the underlying IP ranges – no manual maintenance required.

```
# Example SmartGroup definition for AzureActiveDirectory Service Tag
resource "aviatrix_smart_group" "aca_aad" {
 name = "aca-svctag-aad"
 selector {
  match_expressions {
   external     = "azureips"
   service_name  = "AzureActiveDirectory"
  }
 }
}

# Repeat for: MicrosoftContainerRegistry, AzureFrontDoorFirstParty,
AzureContainerRegistry
```

Operational Safety Properties

Three structural properties ensure the deployment is safe to operate in production environments:

Property	Implementation Detail
Monitor before enforce	Every DCF rule deploys in watch mode by default – same rule, same dataplane, logging only. Promote to enforcement by toggling the Enforce flag. Rollback is the same flag in reverse. The dataplane never reloads. Validating hosted agent runtime health in monitor mode before enforcing eliminates the risk of breaking the runtime.
One VNet at a time	Every deployment step is scoped to the Foundry spoke VNet. A policy regression on this VNet has zero impact on any other spoke in the fabric. Back-out is one action: remove the spoke from the agent subnet. Nothing else is touched.
Policy propagation in ~100ms	DCF rule changes propagate via a kernel-level eBPF map update. The dataplane never reloads. A rule change or full back-out completes in under a second with no service disruption to the Foundry agent runtime.

Known Constraints

Constraint	Workaround / Notes
Hosted Agent trust store injection	Full payload inspection of tool-call egress from the Hosted Agent custom container requires the Aviatrix MITM CA in the container trust store. Or you must upload your own CA to Aviatrix control plane.
Inference traffic not in scope	Inference traffic from the agent to Azure OpenAI routes via private endpoint (RFC 1918, never leaves the VNet) and never traverses the Spoke Gateway. This is correct by design – private endpoint traffic is not the exfiltration surface.
Foundry-managed VNet deployments	Agents deployed on Microsoft-managed VNets (not BYOVNet mode) have no customer ENI and no in-path enforcement point available. Migrate to BYOVNet mode before applying this VCA.

Aviatrix Validated Containment Architecture for Azure AI Foundry Agents removes the unknown from agentic deployment by enforcing policy at the network layer. **Ask your Aviatrix account team** for a guided deployment.

Explore Validated Containment Architectures for other AI platforms.

About Aviatrix

Aviatrix® is pioneering the Cloud Native Security Fabric – the architecture the Containment Era requires. The Cloud Native Security Fabric governs every workload communication path across every cloud, every VPC, every Kubernetes cluster, and every serverless function, from a single policy plane. One rule. Universal propagation. Enforced at the workload, not at a chokepoint. Trusted by more than 500 of the world's leading enterprises. For more information, visit aviatrix.ai.